

The Three-Terminal Interactive Lossy Source Coding Problem

Leonardo Rey Vega, Pablo Piantanida and Alfred O. Hero III

Abstract

The three-node multiterminal lossy source coding problem is investigated. We derive an inner bound to the general rate-distortion region of this problem which is a natural extension of the seminal work by Kaspi'85 on the interactive two-terminal source coding problem. It is shown that this (rather involved) inner bound contains several rate-distortion regions of some relevant source coding settings. In this way, besides the non-trivial extension of the interactive two terminal problem, our results can be seen as a generalization and hence unification of several previous works in the field. Specializing to particular cases we obtain novel rate-distortion regions for several lossy source coding problems. We finish by describing some of the open problems and challenges. However, the general three-node multiterminal lossy source coding problem seems to offer a formidable mathematical complexity.

Index Terms

Multiterminal source coding, Wyner-Ziv, rate-distortion region, Berger-Tung inner bound, interactive lossy source coding, distributed lossy source coding.

The material in this paper was partially published in the IEEE International Symposium on Information Theory, Honolulu, Hawaii, USA, June 29 - July 4, 2014 and in the IEEE International Symposium on Information Theory, Hong-Kong, China, June 14 - June 19, 2015. The work of L. Rey Vega was partially supported by project UBACyT 2002013100751BA. The work of P. Piantanida was partially supported by the FP7 Network of Excellence in Wireless COMMUNICATIONS NEWCOM#. The work of A. Hero was partially supported by a DIGITEO Chair from 2008 to 2013 and by US ARO grant W911NF-15-1-0479.

L. Rey Vega is with the Departments of Electronics (FIUBA) and CSC-CONICET, Buenos Aires, Argentina (e-mail: lrey@fi.uba.ar, cgalar@fi.uba.ar).

P. Piantanida is with the Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, 91192 Gif-sur-Yvette, France (e-mail: pablo.piantanida@supelec.fr).

Alfred O. Hero III is with the Department of Electrical Eng. & CompSci University of Michigan, Ann Arbor, MI, USA (e-mail: hero@umich.edu).

I. INTRODUCTION

A. Motivation and related works

Distributed source coding is an important branch of study in information theory with enormous relevance for the present and future technology. Efficient distributed data compression may be the only way to guarantee acceptable levels of performance when energy and link bandwidth are severely limited as in many real world sensor networks. The distributed data collected by different nodes in a network can be highly correlated and this correlation can be exploited at the application layer, e.g., for target localization and tracking or anomaly detection. In such cases cooperative joint data-compression can achieve a better overall rate-distortion trade-off than independent compression at each node.

Complete answers to the optimal trade-offs between rate and distortion for distributed source coding are scarce and the solution to many problems remain elusive. Two of the most important results in information theory, Slepian-Wolf solution to the distributed lossless source coding problem [1] and Wyner-Ziv [2] single letter solution for the rate-distortion region when side information is available at the decoder provided the kick-off for the study of these important problems. Berger and Tung [3], [4] generalized the Slepian-Wolf problem when lossy reconstructions are required at the decoder. It was shown that the region obtained, although not tight in general, is the optimal one in several special cases [5]–[8] and strictly suboptimal in others [9]. Heegard and Berger [10] considered the Wyner-Ziv problem when the side information at the decoder may be absent or when there are two decoders with degraded side information. Timo *et al* [11] correctly extended the achievable region for many (> 2) decoders. In [12] and the references therein, the complementary delivery problem (closely related to the Heegard-Berger problem) is also studied. The use of interaction in a multiterminal source coding setting has not been so extensively studied as the problems mentioned above. Through the use of multiple rounds of interactive exchanges of information explicit cooperation can take place using distributed/successive refinement source coding. Transmitting “reduced pieces” of information, and constructing an explicit sequential cooperative exchange of information, can be more efficient than transmitting the “total information” in one-shot.

The value of interaction for source coding problems was first recognized by Kaspi in his seminal work [13], where the interactive two-terminal lossy source coding problem was introduced

and solved under the assumption of a finite number of communication rounds. In [14] it is shown that interaction strictly outperforms (in term of sum rate) the Wyner-Ziv rate function. There are also several extensions to the original Kaspi problem. In [15] the interactive source coding problem with a helper is solved when the sources satisfy a certain Markov chain property. In [16]–[18] other interesting cases where interactive cooperation can be beneficial are studied. To the best of our knowledge, a proper generalization of this setting to interactive multiterminal (> 2) lossy source coding has not yet been observed.

B. Main contributions

In this paper, we consider the three-terminal interactive lossy source coding problem presented in Fig. 1. We have a network composed of 3 nodes which can interact through a broadcast rate-limited –error free– channel. Each node measures the realization of a discrete memoryless source (DMS) and is required to reconstruct the sources from the other terminals with a fidelity criterion. Nodes are allowed to interact by interchanging descriptions of their observed sources realizations over a finite number of communication rounds. After the information exchange phase is over, the nodes try to reconstruct the realization of the sources at the other nodes using the recovered descriptions.

The general rate-distortion region seems to pose a formidable mathematical problem which encompass several known open problems. However, several properties of this problem are established in this paper.

General achievable region

We derive a general achievable region by assuming a finite number of rounds. This region is not a trivial extension of Kaspi’s region [13] and the main ideas behind its derivation are the exchange of common and private descriptions between the nodes in the network in order to exploit optimality the different side informations at the different nodes. As in the original Kaspi’s formulation, the key to obtaining the achievable region is the natural cooperation between the nodes induced by the generation of new descriptions based on the past exchanged description. However, in comparison to Kaspi’s 2 node case, the 3 nodes interactions make significant differences in the optimal action of each node at the encoding and decoding procedure in a given round. At each encoding stage, each node need to communicate to two nodes with

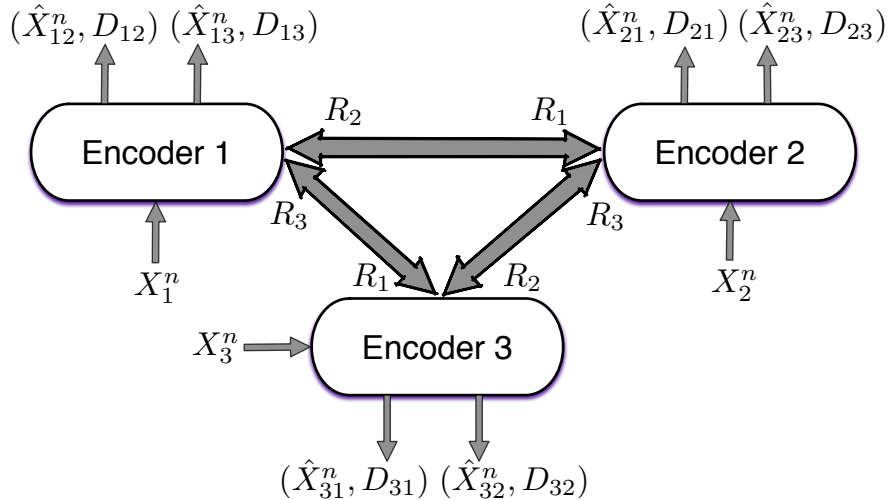


Figure 1: Three-Terminal Interactive Source Coding. There is a single noiseless rate-limited broadcast channel from each terminal to the other two terminals. D_{ij} denotes the average per-letter distortion between the source X_j^n and \hat{X}_{ij}^n measured at the node i for each pair $i \neq j$.

different side information. This is reminiscent of the Heegard-Berger problem [10], [11], whose complete solution is not known, when the side information at the decoders is not degraded. Moreover, the situation is a bit more complex because of the presence of 3-way interaction. This similarity between the Heegard-Berger problem leads us to consider the generation of two sets of messages at each node: common messages destined to all nodes and private messages destined to some restricted sets of nodes. On the other hand, when each node is acting as a decoder the nodes need to recover a set of common and private messages generated at different nodes (i.e. at round l node 3, needs to recover the common descriptions generated at nodes 1 and 2 and the private ones generated also at nodes 1 and 2). This is reminiscent of the Berger-Tung problem [4]–[6], [19] which is also an open problem. Again, the situation is more involved because of the cooperation induced by the multiple rounds of exchanged information. Particularly important is the fact that, in the case of the common descriptions, there is cooperation based on the conditioning on the previous exchanged descriptions in addition to cooperation naturally induced by the encoding-decoding ordering imposed by the network. This explicit cooperation for the exchange of common messages is accomplished through the use of a special binning technique to be explained in Appendix B.

Besides the complexity of the achievable region, we give an inner bound to the rate-distortion region that allows us to recover the two node Kaspi's region. We also recover several previous inner bounds and rate-distortion regions of some well-known cooperative and interactive –as well as non-interactive– lossy source coding problems.

Special cases

As the full problem seems to offer a formidable mathematical complexity, including several special cases which are known to be long-standing open problems, we cannot give a full converse proving the optimality of the general achievable region obtained. However, in Section V we provide a complete answer to the rate-distortion regions of several specific cooperative and interactive source coding problems:

- (1) Two encoders and one decoder subject to lossy/lossless reconstruction constraints without side information (see Fig. 2).
- (2) Two encoders and three decoders subject to lossless/lossy reconstruction constraints with side information (see Fig. 3).
- (3) Two encoders and three decoders subject to lossless/lossy reconstruction constraints, reversal delivery and side information (see Fig. 4).
- (4) Two encoders and three decoders subject to lossy reconstruction constraints with degraded side information (see Fig. 5).
- (5) Three encoders and three decoders subject to lossless/lossy reconstruction constraints with degraded side information (see Fig. 6).

Interestingly enough, we show that for the two last problems, interaction through multiple rounds could be helpful. Whereas for the other three cases, it is shown that a single round of cooperatively exchanged descriptions suffices to achieve optimality. Table I summarizes the characteristics of each of the above mentioned cases.

Next we summarize the contents of the paper. In Section II we formulate the general problem. In Section III we present and discuss the inner bound of the general problem. In Section IV we show how our inner bound contains several results previously obtained in the past. In Section V we present the converse results and their tightness with respect to the inner bound for the special cases mentioned above providing the optimal characterization for them. In Section VI we present a discussion of the obtained results and their limitations and some numerical results

Cases	R_1	R_2	R_3	Constraints at Node 1	Constraints at Node 2	Constraints at Node 3
(1)	$\neq 0$	$\neq 0$	$= 0$	\emptyset (is not reconstructing any source)	\emptyset (is not reconstructing any source)	$\Pr(\hat{X}_{31}^n \neq X_1^n) \leq \epsilon$ $\mathbb{E}[d(\hat{X}_{32}^n, X_2^n)] \leq D_{32}$
(2)	$\neq 0$	$\neq 0$	$= 0$	$\mathbb{E}[d(\hat{X}_{12}^n, X_2^n)] \leq D_{12}$	$\Pr(\hat{X}_{21}^n \neq X_1^n) \leq \epsilon$	$\Pr(\hat{X}_{31}^n \neq X_1^n) \leq \epsilon$ $\mathbb{E}[d(\hat{X}_{32}^n, X_2^n)] \leq D_{32}$
(3)	$\neq 0$	$\neq 0$	$= 0$	$\mathbb{E}[d(\hat{X}_{12}^n, X_2^n)] \leq D_{12}$	$\Pr(\hat{X}_{21}^n \neq X_1^n) \leq \epsilon$	$\Pr(\hat{X}_{32}^n \neq X_2^n) \leq \epsilon$ $\mathbb{E}[d(\hat{X}_{31}^n, X_1^n)] \leq D_{31}$
(4)	$\neq 0$	$\neq 0$	$= 0$	$\mathbb{E}[d(\hat{X}_{12}^n, X_2^n)] \leq D_{12}$	$\mathbb{E}[d(\hat{X}_{21}^n, X_1^n)] \leq D_{21}$	$\mathbb{E}[d(\hat{X}_{31}^n, X_1^n)] \leq D_{31}$ $\mathbb{E}[d(\hat{X}_{32}^n, X_2^n)] \leq D_{32}$
(5)	$\neq 0$	$\neq 0$	$\neq 0$	\emptyset (is not decoding)	$\Pr(\hat{X}_{21}^n \neq X_1^n) \leq \epsilon$ $\mathbb{E}[d(\hat{X}_{23}^n, X_3^n)] \leq D_{23}$	$\Pr(\hat{X}_{31}^n \neq X_1^n) \leq \epsilon$ $\mathbb{E}[d(\hat{X}_{32}^n, X_2^n)] \leq D_{32}$

Table I: Special cases fully characterized in Section V.

concerning the new optimal cases from the previous Section. Finally in Section VII we provide some conclusions. The major mathematical details are relegated to the appendixes.

Notation: We summarize the notation. With x^n and upper-case letters X^n we denote vectors and random vectors of n components, respectively. The i -th component of vector x^n is denoted as x_i . All alphabets are assumed to be finite.

Entropy is denoted by $H(\cdot)$ and mutual information by $I(\cdot; \cdot)$. $H_2(p)$ denotes the entropy associated with a Bernoulli random variable with parameter p . With $h(\cdot)$ we denote differential entropy. Let X, Y and V be three random variables on some alphabets with probability distribution p_{XYV} . When clear from context we will simple denote $p_X(x)$ with $p(x)$. If the probability distribution of random variables X, Y, V satisfies $p(x|yv) = p(x|y)$ for each x, y, v , then they form a Markov chain, which is denoted by $X \text{---} Y \text{---} V$.

The probability of an event \mathcal{A} is denoted by $\Pr\{\mathcal{A}\}$, where the measure used to compute it will be understood from the context. Conditional probability of a set \mathcal{A} with respect to a set \mathcal{B} is denoted as $\Pr\{\mathcal{A}|\mathcal{B}\}$. The set of strong typical sequences associated with random variable X (see appendix A) is denoted by $\mathcal{T}_{[X]^\epsilon}^n$, where $\epsilon > 0$. We simply denote these sets as \mathcal{T}_ϵ^n when clear from the context. The cardinal of set \mathcal{A} is denoted by $\|\mathcal{A}\|$. The complement of a set is denoted by $\bar{\mathcal{A}}$. With $\mathbb{Z}_{\geq \alpha}$ and $\mathbb{R}_{\geq \beta}$ we denote the integers and reals numbers greater than α and

β respectively. $\text{co}\{\mathcal{A}\}$ denotes the convex hull of a set $\mathcal{A} \in \mathbb{R}^N$, where $N \in \mathbb{N}$.

II. PROBLEM FORMULATION

Assume three discrete memoryless sources (DMS's) with alphabets and pmfs given by $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3, p_{X_1 X_2 X_3})$ and arbitrary bounded distortion measures: $d_j : \mathcal{X}_j \times \hat{\mathcal{X}}_j \rightarrow \mathbb{R}_{\geq 0}$, $j \in \mathcal{M} \triangleq \{1, 2, 3\}$ where $\{\hat{\mathcal{X}}_j\}_{j \in \mathcal{M}}$ are finite reconstruction alphabets¹. We consider the problem of characterizing the rate-distortion region of the interactive source coding scenario described in Fig. 1. In this setting, through K rounds of information exchange between the nodes each one of them will attempt to recover a lossy description of the sources that the others nodes observe, e.g., node 1 must reconstruct –while satisfying distortion constraints– the realization of the sources X_2^n and X_3^n observed by nodes 2 and 3. Indeed, this setting can be seen as a generalization of the well-known Kaspi's problem [13].

Definition 1 (K-step interactive source code): A K -step interactive n -length source code, denoted for the network model in Fig. 1, is defined by a sequence of encoder mappings:

$$f_1^l : \mathcal{X}_1^n \times (\mathcal{J}_2^1 \times \mathcal{J}_3^1 \times \cdots \times \mathcal{J}_2^{l-1} \times \mathcal{J}_3^{l-1}) \longrightarrow \mathcal{J}_1^l, \quad (1)$$

$$f_2^l : \mathcal{X}_2^n \times (\mathcal{J}_1^1 \times \mathcal{J}_3^1 \times \cdots \times \mathcal{J}_3^{l-1} \times \mathcal{J}_1^l) \longrightarrow \mathcal{J}_2^l, \quad (2)$$

$$f_3^l : \mathcal{X}_3^n \times (\mathcal{J}_1^1 \times \mathcal{J}_2^1 \times \cdots \times \mathcal{J}_1^l \times \mathcal{J}_2^l) \longrightarrow \mathcal{J}_3^l, \quad (3)$$

with $l \in [1 : K]$ and message sets: $\mathcal{J}_i^l \triangleq \{1, 2, \dots, \mathcal{I}_i^l\}$, $\mathcal{I}_i^l \in \mathbb{Z}_{\geq 0}$, $i \in \mathcal{M}$, and reconstruction mappings:

$$g_{ij} : \mathcal{X}_i^n \times \bigotimes_{m \in \mathcal{M}, m \neq i} (\mathcal{J}_m^1 \times \cdots \times \mathcal{J}_m^K) \longrightarrow \hat{\mathcal{X}}_{ij}^n, \quad i \neq j. \quad (4)$$

The average per-letter distortion and the corresponding distortion levels achieved at the node i with respect to source j are:

$$\mathbb{E} \left[d_j \left(X_j^n, \hat{X}_{ij}^n \right) \right] \leq D_{ij} \quad i, j \in \mathcal{M}, \quad i \neq j \quad (5)$$

with

$$d(x^n, y^n) \equiv \frac{1}{n} \sum_{m=1}^n d(x_m, y_m). \quad (6)$$

¹The problem can be easily generalized to the case in which there are different reconstruction alphabets at the terminals. It can also be shown that all the results are valid if we employ arbitrary bounded joint distortion functions, e.g. at node 1 we use $d(X_2, X_3; \hat{X}_2, \hat{X}_3)$.

In compact form we denote a K -step interactive source coding by $(n, K, \mathcal{F}, \mathcal{G})$ where \mathcal{F} and \mathcal{G} denote the sets of encoders and decoders mappings.

Remark 1: The code definition depends on the node ordering in the encoding procedure. Above we defined the encoding functions $\{f_1^l, f_2^l, f_3^l\}_{l=1}^K$ assuming that in each round node 1 acts first, followed by node 2, and finally by node 3, and the process beginning again at node 1.

Definition 2 (Achievability and rate-distortion region): Consider $\mathbf{R} \triangleq (R_1, R_2, R_3)$ and $\mathbf{D} \triangleq (D_{12}, D_{13}, D_{21}, D_{23}, D_{31}, D_{32})$. The rate vector \mathbf{R} is (\mathbf{D}, K) -achievable if $\forall \epsilon > 0$ there is $n_0(\epsilon, K) \in \mathbb{N}$ such that $\forall n > n_0(\epsilon, K)$ there exists a K -step interactive source code $(n, K, \mathcal{F}, \mathcal{G})$ with rates satisfying:

$$\frac{1}{n} \sum_{l=1}^K \log \|\mathcal{J}_i^l\| \leq R_i + \epsilon, \quad i \in \mathcal{M} \quad (7)$$

and with average per-letter distortions at node i with respect to source j :

$$\mathbb{E} \left[d_j(X_j^n, \hat{X}_{ij}^n) \right] \leq D_{ij} + \epsilon, \quad i, j \in \mathcal{M}, \quad i \neq j, \quad (8)$$

where $\hat{X}_{ij}^n \equiv g_{ij} \left(X_i^n, \bigotimes_{m \in \mathcal{M}, m \neq i} (\mathcal{J}_m^1 \times \cdots \times \mathcal{J}_m^K) \right)$, $i \neq j \in \mathcal{M}$. The rate-distortion region $\mathcal{R}_3(\mathbf{D}, K)$ is defined by:

$$\mathcal{R}_3(\mathbf{D}, K) = \left\{ \mathbf{R} : \mathbf{R} \text{ is } (\mathbf{D}, K)\text{-achievable} \right\} \quad (9)$$

Similarly, the \mathbf{D} -achievable region $\mathcal{R}_3(\mathbf{D})$ is given by $\mathcal{R}_3(\mathbf{D}) = \bigcup_{K=1}^{\infty} \mathcal{R}_3(\mathbf{D}, K)^2$, that is:

$$\mathcal{R}_3(\mathbf{D}) = \left\{ \mathbf{R} : \mathbf{R} \text{ is } (\mathbf{D}, K)\text{-achievable for some } K \in \mathbb{Z}_{\geq 1} \right\}. \quad (10)$$

Remark 2: By definition $\mathcal{R}_3(\mathbf{D}, K)$ is closed and using a time-sharing argument it is easy to show that it is also convex $\forall K \in \mathbb{Z}_{\geq 1}$.

Remark 3: $\mathcal{R}_3(\mathbf{D}, K)$ depends on the node ordering in the encoding procedure. Above we defined the encoding functions $\{f_1^l, f_2^l, f_3^l\}_{l=1}^K$ assuming that in each round node 1 acts first, followed by node 2, and finally by node 3, and the process beginning again at node 1. In this paper we restrict the analysis to the *canonical ordering* $(1 \rightarrow 2 \rightarrow 3)$. However, there are $3! = 6$ different orderings that generally lead to different regions and the (\mathbf{D}, K) -achievable region

²Notice that this limit exists because it is the union of a monotone increasing sequence of sets.

defined above is more explicitly denoted $\mathcal{R}_3(\mathbf{D}, K, \sigma_c)$, where σ_c is the trivial permutation for \mathcal{M} . The correct (\mathbf{D}, K) -achievable region is:

$$\mathcal{R}_3(\mathbf{D}, K) = \bigcup_{\sigma \in \Sigma(\mathcal{M})} \mathcal{R}_3(\mathbf{D}, K, \sigma) \quad (11)$$

where $\Sigma(\mathcal{M})$ contains all the permutations of set \mathcal{M} . The theory presented in this paper for determining $\mathcal{R}_3(\mathbf{D}, K, \sigma_c)$ can be used on the other permutations $\sigma \neq \sigma_c$ to compute (11)³.

III. INNER BOUND ON THE RATE-DISTORTION REGION

In this Section, we provide a general achievable rate-region on the rate-distortion region.

A. Inner bound

We first present a general achievable rate-region where each node at a given round l will generate descriptions destined to the other nodes based on the realization of its own source, the past descriptions generated by a particular node and the descriptions generated at the other nodes and recovered by the node up to the present round. In order to precisely describe the complex rate-region, we need to introduce some definitions. For a set \mathcal{A} , let $\mathcal{C}(\mathcal{A}) = 2^{\mathcal{A}} \setminus \{\mathcal{A}, \emptyset\}$ be the set of all subsets of \mathcal{A} minus \mathcal{A} and the empty set. Denote the auxiliary random variables:

$$U_{i \rightarrow S, l}, \quad S \in \mathcal{C}(\mathcal{M}), \quad i \notin S, \quad l = 1, \dots, K. \quad (12)$$

Auxiliary random variables $\{U_{i \rightarrow S, l}\}$ will be used to denote the descriptions generated at node i and at round l and destined to a set of nodes $S \in \mathcal{C}(\mathcal{M})$ with $i \notin S$. For example, $U_{1 \rightarrow 23, l}$ denote the description generated at node 1 and at round l and destined to nodes 2 and 3. Similarly, $\{U_{1 \rightarrow 2, l}\}$ will be used to denote the descriptions generated at node 1 at round l and destined only to node 2. We define variables:

$\mathcal{W}_{[i, l]} \equiv \text{Common information}^4$ shared by the three nodes available at node i at round l before encoding

³It should be mentioned that this is not the most general setting of the problem. The most general encoding procedure will follow from the definition of the transmission order by a sequence $t_1, t_2, t_3, \dots, t_{\|\mathcal{M}\| \times K}$ with $t_i \in \mathcal{M}$. This will cover even the situation in which the order can be changed in each round. To keep the mathematical presentation simpler we will not consider this more general setting.

⁴Not to be confused with the Wyner's definition of *common information* [20].

$\mathcal{V}_{[S,l,i]} \equiv$ Private information shared by nodes in $S \in \mathcal{C}(\mathcal{M})$ available at node $i \in S$, at round l , before encoding

In precise terms, the quantities introduced above for our problem are defined by:

$$\begin{aligned}\mathcal{W}_{[1,l]} &= \{U_{1 \rightarrow 23,k}, U_{2 \rightarrow 13,k}, U_{3 \rightarrow 12,k}\}_{k=1}^{l-1}, \\ \mathcal{W}_{[2,l]} &= \mathcal{W}_{[1,l]} \cup U_{1 \rightarrow 23,l}, \\ \mathcal{W}_{[3,l]} &= \mathcal{W}_{[2,l]} \cup U_{2 \rightarrow 13,l}, \\ \mathcal{V}_{[12,l,1]} &= \{U_{1 \rightarrow 2,k}, U_{2 \rightarrow 1,k}\}_{k=1}^{l-1}, \quad \mathcal{V}_{[12,l,2]} = \mathcal{V}_{[12,l,1]} \cup U_{1 \rightarrow 2,l}, \\ \mathcal{V}_{[13,l,1]} &= \{U_{1 \rightarrow 3,k}, U_{3 \rightarrow 1,k}\}_{k=1}^{l-1}, \quad \mathcal{V}_{[13,l,3]} = \mathcal{V}_{[13,l,1]} \cup U_{1 \rightarrow 3,l}, \\ \mathcal{V}_{[23,l,2]} &= \{U_{2 \rightarrow 3,k}, U_{3 \rightarrow 2,k}\}_{k=1}^{l-1}, \quad \mathcal{V}_{[23,l,3]} = \mathcal{V}_{[23,l,2]} \cup U_{2 \rightarrow 3,l}.\end{aligned}$$

Before presenting the general inner bound, we provide the basic idea of the *random coding* scheme that achieves the rate-region in Theorem 1 for the case of K communication rounds.

Assume that all codebooks are randomly generated and known to all the nodes before the information exchange begins and consider the encoding ordering given by $1 \rightarrow 2 \rightarrow 3$ so that we begin at round $l = 1$ in node 1. Also, and in order to maintain the explanation simple and to help the reader to grasp the essentials of the coding scheme employed, we will consider that all terminal are able to recover the descriptions generated at other nodes (which will be the case under the conditions in our Theorem 1). From the observation of the source X_1^n , node 1 generates a set of descriptions for each of the other nodes connected to it. In particular it generates a common description to be recovered at nodes 2 and 3 in addition to two private descriptions for node 2 and 3, respectively, generated from a conditional codebook given the common description. Then, node 2 tries to recover the descriptions destined to it (the common description generated at 1 and its corresponding private description), using X_2^n as side information, and generates its own descriptions, based on source X_2^n and the recovered descriptions from node 1. Again, it generates a common description for nodes 1 and 3, a private description for node 3 and another one for node 1. The same process goes on until node 3, which tries to recover jointly the common descriptions generated by node 1 and node 2, and then the private descriptions destined to him by node 1 and 2. Then generates its own descriptions (common and private ones) destined to nodes 1 and 2. Finally, node 1 tries to recover all the descriptions destined to it generated by nodes 2 and 3 in the same way as previously done by node 3. After this, round $l = 1$ is over,

and round $l = 2$ begins with node 1 generating new descriptions using X_1^n , its encoding history (from previous round) and the recovered descriptions from the other nodes.

The process continues in a similar manner until we reach round $l = K$ where node 3 recovers the descriptions from the other nodes and generates its own ones. Node 1 recovers the last descriptions destined to it from nodes 2 and 3 but does not generate new ones. The same holds for node 2 who only recovers the descriptions generated by node 3 and thus terminating the information exchange procedure. Notice that at the end of round K the decoding in node 1 and node 2 can be done simultaneously. This is due to the fact that node 1 is not generating a new description destined to node 2. However, in order to simplify the analysis and notation in the appendix we will consider that the last decoding of node 2 occurs in round $K + 1$ ⁵. After all the exchanges are done, each node recovers an estimate of the other nodes, source realizations by using all the available recovered descriptions from the K previous rounds.

Theorem 1 (Inner bound): Let $\bar{\mathcal{R}}_3(\mathbf{D}, K)$ be the closure of set of all rate tuples satisfying:

$$R_1 = \sum_{l=1}^K (R_{1 \rightarrow 23}^{(l)} + R_{1 \rightarrow 2}^{(l)} + R_{1 \rightarrow 3}^{(l)}) \quad (13)$$

$$R_2 = \sum_{l=1}^K (R_{2 \rightarrow 13}^{(l)} + R_{2 \rightarrow 1}^{(l)} + R_{2 \rightarrow 3}^{(l)}) \quad (14)$$

$$R_3 = \sum_{l=1}^K (R_{3 \rightarrow 12}^{(l)} + R_{3 \rightarrow 1}^{(l)} + R_{3 \rightarrow 2}^{(l)}) \quad (15)$$

$$R_1 + R_2 = \sum_{l=1}^K (R_{1 \rightarrow 23}^{(l)} + R_{2 \rightarrow 13}^{(l)} + R_{1 \rightarrow 3}^{(l)} + R_{2 \rightarrow 3}^{(l)} + R_{1 \rightarrow 2}^{(l)} + R_{2 \rightarrow 1}^{(l)}) \quad (16)$$

$$R_1 + R_3 = \sum_{l=1}^{K+1} (R_{1 \rightarrow 23}^{(l)} + R_{3 \rightarrow 12}^{(l-1)} + R_{1 \rightarrow 2}^{(l)} + R_{3 \rightarrow 2}^{(l-1)} + R_{1 \rightarrow 3}^{(l)} + R_{3 \rightarrow 1}^{(l)}) \quad (17)$$

$$R_2 + R_3 = \sum_{l=1}^K (R_{2 \rightarrow 13}^{(l)} + R_{3 \rightarrow 12}^{(l)} + R_{2 \rightarrow 1}^{(l)} + R_{3 \rightarrow 1}^{(l)} + R_{2 \rightarrow 3}^{(l)} + R_{3 \rightarrow 2}^{(l)}) \quad (18)$$

⁵This is clearly a fictitious round, in the sense that there is not descriptions generation on it. In this way, there is not modification of the final rates achieved by the procedure described if we consider this additional round.

where⁶ for each $l \in [1 : K]$:

$$R_{1 \rightarrow 23}^{(l)} > I \left(X_1; U_{1 \rightarrow 23, l} \middle| X_2 \mathcal{W}_{[1, l]} \mathcal{V}_{[12, l, 1]} \mathcal{V}_{[23, l-1, 3]} \right) \quad (19)$$

$$R_{2 \rightarrow 13}^{(l)} > I \left(X_2; U_{2 \rightarrow 13, l} \middle| X_3 \mathcal{W}_{[2, l]} \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]} \right) \quad (20)$$

$$R_{3 \rightarrow 12}^{(l)} > I \left(X_3; U_{3 \rightarrow 12, l} \middle| X_1 \mathcal{W}_{[3, l]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l, 3]} \right) \quad (21)$$

$$R_{1 \rightarrow 23}^{(l)} + R_{2 \rightarrow 13}^{(l)} > I \left(X_1 X_2; U_{1 \rightarrow 23, l} U_{2 \rightarrow 13, l} \middle| X_3 \mathcal{W}_{[1, l]} \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]} \right) \quad (22)$$

$$R_{2 \rightarrow 13}^{(l)} + R_{3 \rightarrow 12}^{(l)} > I \left(X_2 X_3; U_{2 \rightarrow 13, l} U_{3 \rightarrow 12, l} \middle| X_1 \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l, 3]} \right) \quad (23)$$

$$R_{1 \rightarrow 23}^{(l)} + R_{3 \rightarrow 12}^{(l-1)} > I \left(X_1 X_3; U_{1 \rightarrow 23, l} U_{3 \rightarrow 12, l-1} \middle| X_2 \mathcal{W}_{[3, l-1]} \mathcal{V}_{[12, l, 1]} \mathcal{V}_{[23, l-1, 3]} \right) \quad (24)$$

$$R_{3 \rightarrow 2}^{(l-1)} > I \left(X_3; U_{3 \rightarrow 2, l-1} \middle| X_2 \mathcal{W}_{[2, l]} \mathcal{V}_{[23, l-1, 3]} \mathcal{V}_{[12, l, 2]} \right) \quad (25)$$

$$R_{1 \rightarrow 2}^{(l)} > I \left(X_1; U_{1 \rightarrow 2, l} \middle| X_2 \mathcal{W}_{[2, l]} \mathcal{V}_{[23, l, 2]} \mathcal{V}_{[12, l, 1]} \right) \quad (26)$$

$$R_{1 \rightarrow 2}^{(l)} + R_{3 \rightarrow 2}^{(l-1)} > I \left(X_1 X_3; U_{1 \rightarrow 2, l} U_{3 \rightarrow 2, l-1} \middle| X_2 \mathcal{W}_{[2, l]} \mathcal{V}_{[23, l-1, 3]} \mathcal{V}_{[12, l, 1]} \right) \quad (27)$$

$$R_{1 \rightarrow 3}^{(l)} > I \left(X_1; U_{1 \rightarrow 3, l} \middle| X_3 \mathcal{W}_{[3, l]} \mathcal{V}_{[23, l, 3]} \mathcal{V}_{[13, l, 1]} \right) \quad (28)$$

$$R_{2 \rightarrow 3}^{(l)} > I \left(X_2; U_{2 \rightarrow 3, l} \middle| X_3 \mathcal{W}_{[3, l]} \mathcal{V}_{[23, l, 2]} \mathcal{V}_{[13, l, 3]} \right) \quad (29)$$

$$R_{1 \rightarrow 3}^{(l)} + R_{2 \rightarrow 3}^{(l)} > I \left(X_1 X_2; U_{1 \rightarrow 3, l} U_{2 \rightarrow 3, l} \middle| X_3 \mathcal{W}_{[3, l]} \mathcal{V}_{[23, l, 2]} \mathcal{V}_{[13, l, 1]} \right) \quad (30)$$

$$R_{2 \rightarrow 1}^{(l)} > I \left(X_2; U_{2 \rightarrow 1, l} \middle| X_1 \mathcal{W}_{[1, l+1]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l+1, 1]} \right) \quad (31)$$

$$R_{3 \rightarrow 1}^{(l)} > I \left(X_3; U_{3 \rightarrow 1, l} \middle| X_1 \mathcal{W}_{[1, l+1]} \mathcal{V}_{[12, l+1, 1]} \mathcal{V}_{[13, l, 3]} \right) \quad (32)$$

$$R_{2 \rightarrow 1}^{(l)} + R_{3 \rightarrow 1}^{(l)} > I \left(X_2 X_3; U_{2 \rightarrow 1, l} U_{3 \rightarrow 1, l} \middle| X_1 \mathcal{W}_{[1, l+1]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l, 3]} \right) \quad (33)$$

with $R_{i \rightarrow S}^{(0)} = R_{i \rightarrow S}^{(K+1)} = 0$ and $U_{i \rightarrow S, 0} = U_{i \rightarrow S, K+1} = \emptyset$ for $S \in \mathcal{C}(\mathcal{M})$ and $i \notin S$.

With these definitions the rate-distortion region satisfies⁷:

$$\bigcup_{p \in \mathcal{P}(\mathbf{D}, K)} \bar{\mathcal{R}}_3(\mathbf{D}, K) \subseteq \mathcal{R}_3(\mathbf{D}, K) , \quad (34)$$

⁶Notice that these definitions are motivated by the fact that at round 1, node 2 only recovers the descriptions generated by node 1 and at round $K + 1$ only recovers what node 3 already generated at round K .

⁷It is straightforward to show that the LHS of equation (34) is convex, which implies that the convex hull operation is not needed.

where $\mathcal{P}(\mathbf{D}, K)$ denotes the set of all joint probability measures associated with the following Markov chains for every $l \in [1 : K]$:

- 1) $U_{1 \rightarrow 23, l} \ominus (X_1, \mathcal{W}_{[1, l]}) \ominus (X_2, X_3, \mathcal{V}_{[12, l, 1]}, \mathcal{V}_{[13, l, 1]}, \mathcal{V}_{[23, l, 2]})$,
- 2) $U_{1 \rightarrow 2, l} \ominus (X_1, \mathcal{W}_{[2, l]}, \mathcal{V}_{[12, l, 1]}) \ominus (X_2, X_3, \mathcal{V}_{[13, l, 1]}, \mathcal{V}_{[23, l, 2]})$,
- 3) $U_{1 \rightarrow 3, l} \ominus (X_1, \mathcal{W}_{[2, l]}, \mathcal{V}_{[13, l, 1]}) \ominus (X_2, X_3, \mathcal{V}_{[12, l, 2]}, \mathcal{V}_{[23, l, 2]})$,
- 4) $U_{2 \rightarrow 13, l} \ominus (X_2, \mathcal{W}_{[2, l]}) \ominus (X_1, X_3, \mathcal{V}_{[12, l, 2]}, \mathcal{V}_{[13, l, 3]}, \mathcal{V}_{[23, l, 2]})$,
- 5) $U_{2 \rightarrow 1, l} \ominus (X_2, \mathcal{W}_{[3, l]}, \mathcal{V}_{[12, l, 2]}) \ominus (X_1, X_3, \mathcal{V}_{[13, l, 3]}, \mathcal{V}_{[23, l, 2]})$,
- 6) $U_{2 \rightarrow 3, l} \ominus (X_2, \mathcal{W}_{[3, l]}, \mathcal{V}_{[23, l, 2]}) \ominus (X_1, X_3, \mathcal{V}_{[12, l+1, 1]}, \mathcal{V}_{[13, l, 3]})$,
- 7) $U_{3 \rightarrow 12, l} \ominus (X_3, \mathcal{W}_{[3, l]}) \ominus (X_1, X_2, \mathcal{V}_{[12, l+1, 1]}, \mathcal{V}_{[13, l, 3]}, \mathcal{V}_{[23, l, 3]})$,
- 8) $U_{3 \rightarrow 1, l} \ominus (X_3, \mathcal{W}_{[1, l+1]}, \mathcal{V}_{[13, l, 3]}) \ominus (X_1, X_2, \mathcal{V}_{[12, l+1, 1]}, \mathcal{V}_{[23, l, 3]})$,
- 9) $U_{3 \rightarrow 2, l} \ominus (X_3, \mathcal{W}_{[1, l+1]}, \mathcal{V}_{[23, l, 3]}) \ominus 1(X_1, X_2, \mathcal{V}_{[12, l+1, 1]}, \mathcal{V}_{[13, l+1, 1]})$,

and such that there exist reconstruction mappings:

$$g_{ij}(X_i, \mathcal{V}_{[ij, K+1, 1]}, \mathcal{W}_{[1, K+1]}) = \hat{X}_{ij} \quad (35)$$

with $\mathbb{E}[d_j(X_j, \hat{X}_{ij})] \leq D_{ij}$ for each $i, j \in \mathcal{M}$ and $i \neq j$.

The proof of this theorem is relegated to Appendix C and relies on the auxiliary results presented in Appendix A and the theorem on the cooperative Berger-Tung problem with side information presented in Appendix B.

Remark 4: It is worth mentioning here that our coding scheme is constrained to use *successive decoding*, i.e., by recovering first the coding layer of common descriptions and then coding layer of private descriptions (at each coding layer each node employ *joint-decoding*). Obviously, this is a sub-optimum procedure since the best scheme would be to use *joint decoding* where both common and private informations can be jointly recovered. However, the analysis of this scheme is much more involved. The associated achievable rate region involves a large number of equations that combine rates belonging to private and common messages from different nodes. Also, several mutual information terms in each of these rate equations cannot be combined, leading to a proliferation of many equations that offer little insight to the problem.

Remark 5: The idea behind our derivation of the achievable region can be extended to any number $M (> 3)$ of nodes in the network. This can be accomplished by generating a greater number of superimposed coding layers. First a layer of codes that generates descriptions destined

to be decoded by all nodes. The next layer corresponding to all subsets of size $M - 1$, etc, until we reach the final layer composed by codes that generate private descriptions for each of nodes. Again, *successive decoding* is used at the nodes to recover the descriptions in these layers destined to them. Of course, the number of required descriptions will increase with the number of nodes as well as the obtained rate-distortion region.

Remark 6: It is interesting to compare the main ideas of our scheme with those of Kaspi [13]. The main idea in [13] is to have a single coding *tree* shared by the two nodes. Each *leaf* in the coding tree is codeword generated either at node 1 or 2. At a given round each node knows (assuming no errors at the encoding and decoding procedures) the path followed in the tree. For example, at round l , node 1, using the knowledge of the path until round l and its source realization generate a leaf (from a set of possible ones) using joint typicality encoding and binning. Node 2, using the same path known at node 1 and its source realization, uses joint typicality decoding to estimate the leaf generate at node 1. If there is no error at these encoding and decoding steps, the previous path is updated with the new leaf and both -node 1 and 2- know the updated path. Node 2 repeats the procedure. This is done until round K where the final path is known at both nodes and used to reconstruct the desired sources.

In the case of three nodes the situation is more involved. At a given round, the encoder at an arbitrary node is seeing two decoders with different side information⁸. In order to simplify the explanation consider that we are at round l in the encoder 1, and that the listening nodes are nodes 2 and 3. This situation forces node 1 to encode two sets of descriptions: one common for the other two nodes and a set of private ones associated with each of the listening nodes 2 and 3. Following the ideas of Kaspi, it is then natural to consider three different coding trees followed by node 1. One coding tree has leaves that are the common descriptions generated and shared by all the nodes in the network. The second tree is composed by leaves that are the private descriptions generated and shared with node 2. The third tree is composed by leaves that are the private descriptions generated and shared with node 3. As the private descriptions *refine* the common ones, depending on the quality of the side information of the node that is the intended recipient, it is clear that descriptions are correlated. For example, the private description destined to node 2, should depend not only on the past private descriptions generated and shared

⁸Because at each node the source realizations are different, and the recovered previous descriptions can also be different.

by nodes 1 and 2, but also on the common descriptions generated at all previous rounds in all the nodes and on the common description generated at the present round in node 1. Something similar happens for the private description destined to node 3. It is clear that as the common descriptions are to be recovered by all the nodes in the network, they can only be conditioned with respect to the past common descriptions generated at previous rounds and with respect to the common descriptions generated at the present round by a node who acted before (i.e. at round l node 1 acts before node 2). The private descriptions, as they are only required to be recovered at some set of nodes, can be generated conditioned on the past exchanged common descriptions and the past private descriptions generated and recovered in the corresponding set of nodes (i.e., the private descriptions exchanged between nodes 1 and 2 at round l , can only be generated conditioned on the past common descriptions generated at nodes 1, 2 and 3 and on the past private descriptions exchanged only between 1 and 2).

We can see clearly that there are basically four paths to be cooperatively followed in the network:

- One path of common descriptions shared by nodes 1, 2 and 3.
- One path of private descriptions shared by nodes 1 and 2.
- One path of private descriptions shared by nodes 1 and 3.
- One path of private descriptions shared by nodes 2 and 3.

It is also clear that each node only follows three of these paths simultaneously. The exchange of common descriptions deserves special mention. Consider the case at round l in node 3. This node needs to recover the common descriptions generated at nodes 1 and 2. But at the moment node 2 generated its own common description, it also recovered the common one generated at node 1. This allows for a natural explicit cooperation between nodes 1 and 2 in order to help node 3 to recover both descriptions. Clearly, this is not the case for private descriptions from nodes 1 and 2 to be recovered at node 3. Node 2 does not recover the private description from node 1 to 3 and cannot generate an explicit collaboration to help node 3 to recover both private descriptions. Note, however, that as both private descriptions will be dependent on previous common descriptions an implicit collaboration (intrinsic to the code generation) is also in force. In appendix B we consider the problem (not in the interactive setting) of generating the explicit cooperation for the common descriptions through the use of what we call a *super-binning* procedure, in order

to use the results for our interactive three-node problem.

IV. KNOWN CASES AND RELATED WORK

Several inner bounds and rate-distortion regions on multiterminal source coding problems can be derived by specializing the inner bound (34). Below we summarize only a few of them.

1) *Distributed source coding with side information* [4], [19]: Consider the distributed source coding problem where two nodes encode separately sources X_1^n and X_2^n to rates (R_1, R_2) and a decoder by using side information X_3^n must reconstruct both sources with average distortion less than D_1 and D_2 , respectively. By considering only one-round/one-way information exchange from nodes 1 and 2 (the encoders) to node 3 (the decoder), the results in [4], [19] can be recovered as a special case of the inner bound (34). Specifically, we set:

$$U_{1 \rightarrow 23,l} = U_{2 \rightarrow 13,l} = U_{3 \rightarrow 12,l} = U_{1 \rightarrow 2,l} = U_{2 \rightarrow 1,l} = U_{3 \rightarrow 1,l} = U_{3 \rightarrow 2,l} = \emptyset, \quad \forall l$$

$$U_{1 \rightarrow 3,l} = U_{2 \rightarrow 3,l} = \emptyset, \quad \forall l > 1.$$

In this case, the Markov chains of Theorem 1 reduce to:

$$U_{1 \rightarrow 3,1} \oplus X_1 \oplus (X_2, X_3, U_{2 \rightarrow 3,1}), \quad (36)$$

$$U_{2 \rightarrow 3,1} \oplus X_2 \oplus (X_1, X_3, U_{1 \rightarrow 3,1}), \quad (37)$$

and thus the inner bound from Theorem 1 recovers the results in [19]

$$R_1 > I(X_1; U_{1 \rightarrow 3,1} | X_3 U_{2 \rightarrow 3,1}), \quad (38)$$

$$R_2 > I(X_2; U_{2 \rightarrow 3,1} | X_3 U_{1 \rightarrow 3,1}), \quad (39)$$

$$R_1 + R_2 > I(X_1 X_2; U_{1 \rightarrow 3,1} U_{2 \rightarrow 3,1} | X_3). \quad (40)$$

2) *Source coding with side information at 2-decoders* [10], [11]: Consider the setting where one encoder X_1^n transmits descriptions to two decoders with different side informations (X_2^n, X_3^n) and distortion requirements D_2 and D_3 . Again we consider only one way/round information exchange from node 1 (the encoder) to nodes 2 and 3 (the decoders).

In this case, we set:

$$U_{2 \rightarrow 13,l} = U_{3 \rightarrow 12,l} = U_{2 \rightarrow 1,l} = U_{3 \rightarrow 1,l} = U_{3 \rightarrow 2,l} = U_{2 \rightarrow 3,l} = \emptyset, \quad \forall l$$

$$U_{1 \rightarrow 23,l} = U_{1 \rightarrow 23,l} = U_{1 \rightarrow 2,l} = U_{1 \rightarrow 3,l} = \emptyset, \quad \forall l > 1.$$

The above Markov chains imply

$$(U_{1 \rightarrow 23,1}, U_{1 \rightarrow 2,1}, U_{1 \rightarrow 3,1}) \multimap X_1 \multimap (X_2, X_3) \quad (41)$$

and thus the inner bound from Theorem 1 reduces to the results in [10], [11]

$$\begin{aligned} R_1 > \max \{ & I(X_1; U_{1 \rightarrow 23,1} | X_2), I(X_1; U_{1 \rightarrow 23,1} | X_3) \} \\ & + I(X_1; U_{1 \rightarrow 2,1} | X_2 U_{1 \rightarrow 23,1}) + I(X_1; U_{1 \rightarrow 3,1} | X_3 U_{1 \rightarrow 23,1}) . \end{aligned} \quad (42)$$

3) *Two terminal interactive source coding [13]:* Our inner bound (34) is basically the generalization of the two terminal problem to the three-terminal setting. Assume only two encoders-decoders X_1^n and X_2^n which must reconstruct the other terminal source 3 with distortion constraints D_1 and D_2 , and after K rounds of information exchange. Let us set:

$$\begin{aligned} U_{1 \rightarrow 23,l} &= U_{2 \rightarrow 13,l} = U_{3 \rightarrow 12,l} = U_{1 \rightarrow 3,l} = U_{3 \rightarrow 1,l} = U_{2 \rightarrow 3,l} = U_{3 \rightarrow 2,l} = \emptyset, \quad \forall l \\ X_3 &= \emptyset . \end{aligned}$$

The Markov chains become

$$U_{1 \rightarrow 2,l} \multimap (X_1, \mathcal{V}_{[12,l,1]}) \multimap X_2, \quad (43)$$

$$U_{2 \rightarrow 1,l} \multimap (X_2, \mathcal{V}_{[12,l,2]}) \multimap X_2, \quad (44)$$

for $l \in [1 : K]$ and thus the inner bound from Theorem 1 permit us to obtain the results in [13]

$$R_1 > I(X_1; \mathcal{V}_{[12,K+1,1]} | X_2), \quad (45)$$

$$R_2 > I(X_2; \mathcal{V}_{[12,K+1,2]} | X_1). \quad (46)$$

4) *Two terminal interactive source coding with a helper [15]:* Consider now two encoders/decoders, namely X_2^n and X_3^n , that must reconstruct the other terminal source with distortion constraints D_2 and D_3 , respectively, using K communication rounds. Assume also that another encoder X_1^n provides both nodes (2, 3) with a common description before beginning the information exchange and then remains silent. Such common description can be exploited as coded side information. Let us set:

$$\begin{aligned} U_{2 \rightarrow 13,l} &= U_{3 \rightarrow 12,l} = U_{1 \rightarrow 3,l} = U_{1 \rightarrow 2,l} = U_{1 \rightarrow 3,l} = U_{2 \rightarrow 1,l} = U_{3 \rightarrow 1,l} = \emptyset, \quad \forall l \\ U_{1 \rightarrow 23,l} &= \emptyset, \quad \forall l > 1 . \end{aligned}$$

The Markov chains reduce to:

$$U_{1 \rightarrow 23,1} \text{---} X_1 \text{---} (X_2, X_3) , \quad (47)$$

$$U_{2 \rightarrow 3,l} \text{---} (X_2, U_{1 \rightarrow 23,1}, \mathcal{V}_{[23,l,2]}) \text{---} (X_1, X_3) , \quad (48)$$

$$U_{3 \rightarrow 2,l} \text{---} (X_3, U_{1 \rightarrow 23,1}, \mathcal{V}_{[23,l,3]}) \text{---} (X_1, X_2) . \quad (49)$$

An inner bound to the rate-distortion region for this problem reduces to (using the rate equations in our Theorem 1)

$$R_1 > \max \{ I(X_1; U_{1 \rightarrow 23,1} | X_2), I(X_1; U_{1 \rightarrow 23,1} | X_3) \} , \quad (50)$$

$$R_2 > I(X_2; \mathcal{V}_{[23,K+1,2]} | X_3 U_{1 \rightarrow 23,1}) , \quad (51)$$

$$R_3 > I(X_3; \mathcal{V}_{[23,K+1,2]} | X_2 U_{1 \rightarrow 23,1}) . \quad (52)$$

This region contains as a special case the region in [15]. In that paper it is further assumed (in order to have a converse result) that $X_1 \text{---} X_3 \text{---} X_2$. Then, the value of R_1 satisfies $R_1 > I(X_1; U_{1 \rightarrow 23,1} | X_2)$. Obviously, with the same extra Markov chain we obtain the same limiting value for R_1 and the above region is the rate-distortion region.

V. NEW RESULTS ON INTERACTIVE AND COOPERATIVE SOURCE CODING

A. Two encoders and one decoder subject to lossy/lossless reconstruction constraints without side information

Consider now the problem described in Fig. 2 where encoder 1 wishes to communicate the source X_1^n to node 3 in a *lossless* manner while encoder 2 wishes to send a *lossy* description of the source X_2^n to node 3 with distortion constraint D_{31} . To achieve this, the encoders use K communication rounds. This problem can be seen as the cooperating encoders version of the well-known Berger-Yeung [5] problem.

Theorem 2: The rate-distortion region of the setting described in Fig. 8 is given by the union over all joint probability measures $p_{X_1 X_2 U_{2 \rightarrow 13}}$ such that there exists a reconstruction mapping:

$$g_{32}(X_1, U_{2 \rightarrow 13}) = \hat{X}_{32} \quad \text{with} \quad \mathbb{E} \left[d(X_2, \hat{X}_{32}) \right] \leq D_{32} , \quad (53)$$

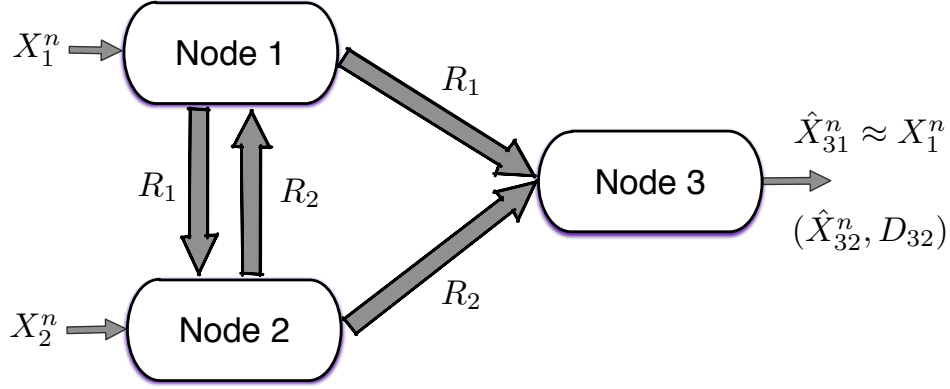


Figure 2: Two encoders and one decoder subject to lossy/lossless reconstruction constraints without side information.

of the set of all tuples satisfying:

$$R_1 \geq H(X_1|X_2) , \quad (54)$$

$$R_2 \geq I(X_2; U_{2 \rightarrow 13} | X_1) , \quad (55)$$

$$R_1 + R_2 \geq H(X_1) + I(X_2; U_{2 \rightarrow 13} | X_1) . \quad (56)$$

The auxiliary random variable $U_{2 \rightarrow 13}$ has a cardinality bound of $\|\mathcal{U}_{2 \rightarrow 13}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_2\| + 1$.

Remark 7: It is worth emphasizing that the rate-distortion region in Theorem 2 outperforms the non-cooperative rate-distortion region first derived in [5]. This is due to two facts: the conditional entropy given in the rate constraint (54) which is strictly smaller than the entropy $H(X_1)$ present in the rate-region in [5], and the fact that the random description $U_{2 \rightarrow 13}$ may be arbitrarily dependent on both sources (X_1, X_2) which is not the case without cooperation [5]. Therefore, cooperation between encoders 1 and 2 reduces the rate needed to communicate the source X_1 while increasing the optimization set of all admissible source descriptions.

Remark 8: Notice that the rate-distortion region in Theorem 2 is achievable with a single round of interactions $K = 1$, which implies that multiple rounds do not improve the rate-distortion region in this case. This holds because of the fact that node 3 reconstruct X_1 in a lossless fashion.

Remark 9: Although in the considered setting of Fig. 8 node 1 is not supposed to decode neither a lossy description nor the complete source X_2^n , if nodes 1 and 3 wish to recover the same

descriptions the optimal rate-region remains the same as given in Theorem 2. The only difference relies on the fact that node 1 is now able to find a function $g_{12}(X_1, U_{2 \rightarrow 13}) = \hat{X}_{12}$ which must satisfy an additional distortion constraint $\mathbb{E} \left[d(X_2, \hat{X}_{12}) \right] \leq D_{12}$. In order to show this, it is enough to check that in the converse proof given below the specific choice of the auxiliary random variable already allows node 1 to recover a general function $\hat{X}_{12[t]} = g_{12}(X_{1[t]}, U_{2 \rightarrow 13[t]})$ for each time $t \in \{1, \dots, n\}$.

Proof: The direct part of the proof simply follows by choosing:

$$\begin{aligned} U_{3 \rightarrow 12, l} &= U_{1 \rightarrow 3, l} = U_{1 \rightarrow 2, l} = U_{2 \rightarrow 1, l} = U_{2 \rightarrow 3, l} = U_{3 \rightarrow 1, l} = U_{3 \rightarrow 2, l} = \emptyset, \quad \forall l \\ U_{1 \rightarrow 23, 1} &= X_1, \quad U_{1 \rightarrow 23, l} = U_{2 \rightarrow 13, l} = \emptyset \quad \forall l > 1, \end{aligned}$$

and thus the rate-distortion region (34) reduces to the desired region in Theorem 2 where for simplicity we dropped the round index. We now proceed to the proof of the converse.

If a pair of rates (R_1, R_2) and distortion D_{32} are admissible for the K -steps interactive cooperative distributed source coding setting described in Fig. 8, then for all $\varepsilon > 0$ there exists $n_0(\varepsilon, K)$, such that $\forall n > n_0(\varepsilon, K)$ there exists a K -steps interactive source code $(n, K, \mathcal{F}, \mathcal{G})$ with intermediate rates satisfying:

$$\frac{1}{n} \sum_{l=1}^K \log \|\mathcal{J}_i^l\| \leq R_i + \varepsilon, \quad i \in \{1, 2\} \quad (57)$$

and with average per-letter distortions with respect to the source 2 and perfect reconstruction with respect to the source 1 at node 3:

$$\mathbb{E} \left[d(X_2^n, \hat{X}_{32}^n) \right] \leq D_{32} + \varepsilon, \quad (58)$$

$$\Pr \left(X_1^n \neq \hat{X}_{31}^n \right) \leq \varepsilon, \quad (59)$$

where

$$\hat{X}_{32}^n \equiv g_{32} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]} \right), \quad \hat{X}_{31}^n \equiv g_{31} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]} \right). \quad (60)$$

For each $t \in \{1, \dots, n\}$, define random variables $U_{2 \rightarrow 13[t]}$ as follows:

$$U_{2 \rightarrow 13[t]} \triangleq \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_{1[1:t-1]}, X_{1[t+1:n]} \right). \quad (61)$$

By the condition (59) which says that $\Pr \left(X_1^n \neq \hat{X}_{31}^n \right) \leq \varepsilon$ and Fano's inequality [21], we have

$$H(X_1^n | \hat{X}_{31}^n) \leq \Pr \left(X_1^n \neq \hat{X}_{31}^n \right) \log_2(\|\mathcal{X}_1^n\| - 1) + H_2 \left(\Pr(X_1^n \neq \hat{X}_{31}^n) \right) \triangleq n\epsilon_n, \quad (62)$$

where $\epsilon_n(\varepsilon) \rightarrow 0$ provided that $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$.

1) *Rate at node 1:* For the first rate, we have

$$n(R_1 + \varepsilon) \geq H\left(\mathcal{J}_1^{[1:K]}\right) \quad (63)$$

$$\geq I\left(\mathcal{J}_1^{[1:K]}; X_1^n | X_2^n\right) \quad (64)$$

$$= nH(X_1 | X_2) - H\left(X_1^n | X_2^n \mathcal{J}_1^{[1:K]}\right) \quad (65)$$

$$\stackrel{(a)}{=} nH(X_1 | X_2) - H\left(X_1^n | X_2^n \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}\right) \quad (66)$$

$$\stackrel{(b)}{\geq} nH(X_1 | X_2) - H(X_1^n | \hat{X}_{31}^n) \quad (67)$$

$$\stackrel{(b)}{\geq} nH(X_1 | X_2) - n\epsilon_n, \quad (68)$$

where

- step (a) follows from the fact that by definition of the code the sequence $\mathcal{J}_2^{[1:K]}$ is a function of the source X_2^n and the vector of messages $\mathcal{J}_1^{[1:K]}$,
- step (b) follows from the code assumption that guarantees the existence of a reconstruction function $\hat{X}_{31}^n \equiv g_{31}\left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}\right)$,
- step (c) follows from Fano's inequality in (62).

2) *Rate at node 2:* For the second rate, we have

$$n(R_2 + \varepsilon) \geq H(\mathcal{J}_2^{[1:K]}) \quad (69)$$

$$\geq I(\mathcal{J}_2^{[1:K]}; X_1^n X_2^n) \quad (70)$$

$$= I(\mathcal{J}_2^{[1:K]}; X_1^n) + I(\mathcal{J}_2^{[1:K]}; X_2^n | X_1^n) \quad (71)$$

$$\stackrel{(a)}{\geq} I(\mathcal{J}_2^{[1:K]}; X_1^n) + \sum_{t=1}^n I(\mathcal{J}_2^{[1:K]}; X_{2[t]} | X_{1[t]} X_{1[t+1:n]} X_{1[1:t-1]} X_{2[1:t-1]}) \quad (72)$$

$$\stackrel{(b)}{=} I(\mathcal{J}_2^{[1:K]}; X_1^n) + \sum_{t=1}^n I(\mathcal{J}_2^{[1:K]} X_{1[t+1:n]} X_{1[1:t-1]} X_{2[1:t-1]}; X_{2[t]} | X_{1[t]}) \quad (73)$$

$$\stackrel{(c)}{\geq} I(\mathcal{J}_2^{[1:K]}; X_1^n) + \sum_{t=1}^n I(U_{2 \rightarrow 13[t]}; X_{2[t]} | X_{1[t]}) \quad (74)$$

$$\stackrel{(d)}{=} I(\mathcal{J}_2^{[1:K]}; X_1^n) + \sum_{t=1}^n I(U_{2 \rightarrow 13[Q]}; X_{2[Q]} | X_{1[Q]}, Q = t) \quad (75)$$

$$\stackrel{(e)}{=} I(\mathcal{J}_2^{[1:K]}; X_1^n) + nI(U_{2 \rightarrow 13[Q]}; X_{2[Q]} | X_{1[Q]}, Q) \quad (76)$$

$$\stackrel{(f)}{\geq} I(\mathcal{J}_2^{[1:K]}; X_1^n) + nI(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \quad (77)$$

$$\stackrel{(g)}{\geq} nI(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1), \quad (78)$$

where

- step (a) follows from the chain rule for conditional mutual information and non-negativity of mutual information,
- step (b) follows from the memoryless property across time of the sources (X_1^n, X_2^n) ,
- step (c) follows from the non-negativity of mutual information and definitions (61),
- step (d) follows from the use of a time sharing random variable Q uniformly distributed over the set $\{1, \dots, n\}$,
- step (e) follows from the definition of the conditional mutual information,
- step (f) follows by letting a new random variable $\tilde{U}_{2 \rightarrow 13} \triangleq (U_{2 \rightarrow 13[Q]}, Q)$,
- step (g) follows from the non-negativity of mutual information.

3) *Sum-rate of nodes 1 and 2:* For the sum-rate, we have

$$n(R_1 + R_2 + 2\varepsilon) \geq H(\mathcal{J}_1^{[1:K]}) + n(R_2 + \varepsilon) \quad (79)$$

$$\stackrel{(a)}{\geq} H(\mathcal{J}_1^{[1:K]}) + I(\mathcal{J}_2^{[1:K]}; X_1^n) + nI(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \quad (80)$$

$$\begin{aligned} &= H(\mathcal{J}_1^{[1:K]} | \mathcal{J}_2^{[1:K]}) + I(\mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]}) \\ &+ I(\mathcal{J}_2^{[1:K]}; X_1^n) + nI(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \end{aligned} \quad (81)$$

$$\begin{aligned} &\geq I(\mathcal{J}_1^{[1:K]}; X_1^n | \mathcal{J}_2^{[1:K]}) + I(\mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]}) \\ &+ I(\mathcal{J}_2^{[1:K]}; X_1^n) + nI(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \end{aligned} \quad (82)$$

$$= I(\mathcal{J}_1^{[1:K]}; X_1^n) + I(X_1^n \mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]}) + nI(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \quad (83)$$

$$\begin{aligned} &\stackrel{(b)}{=} n \left[H(X_1) + I(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \right] \\ &- H(X_1^n | \mathcal{J}_1^{[1:K]}) + I(X_1^n \mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]}) \end{aligned} \quad (84)$$

$$\stackrel{(c)}{\geq} n \left[H(X_1) + I(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \right] - H(X_1^n | \mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]}) \quad (85)$$

$$\stackrel{(d)}{\geq} n \left[H(X_1) + I(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) \right] - H(X_1^n | \hat{X}_{31}^n) \quad (86)$$

$$\stackrel{(e)}{\geq} n \left[H(X_1) + I(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1) - \epsilon_n \right], \quad (87)$$

where

- step (a) follows from inequality (77),
- step (b) follows from the memoryless property across time of the source X_1^n ,
- step (c) follows from non-negativity of mutual information,
- step (d) follows from the code assumption that guarantees the existence of reconstruction function $\hat{X}_{31}^n \equiv g_{31}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]})$ and from the fact that unconditioning increases entropy,
- step (e) from Fano's inequality in (62).

4) *Distortion at node 3:* Node 3 reconstructs lossless $\hat{X}_{31}^n \equiv g_{31}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]})$ and lossy $\hat{X}_{32}^n \equiv g_{32}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]})$. For each $t \in \{1, \dots, n\}$, define a function $\hat{X}_{32[t]}$ as being the t -th coordinate of this estimate:

$$\hat{X}_{32[t]}(U_{2 \rightarrow 13[t]}) \triangleq g_{32[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}) . \quad (88)$$

The component-wise mean distortion thus verifies

$$D_{32} + \varepsilon \geq \mathbb{E} \left[d \left(X_2, g_{31}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}) \right) \right] \quad (89)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[t]}, \hat{X}_{32[t]}(U_{2 \rightarrow 13[t]}) \right) \right] \quad (90)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}) \right) \mid Q = t \right] \quad (91)$$

$$= \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}) \right) \right] \quad (92)$$

$$= \mathbb{E} \left[d \left(X_2, \tilde{X}_{32}(\tilde{U}_{2 \rightarrow 13}) \right) \right], \quad (93)$$

where we defined function \tilde{X}_{32} by

$$\tilde{X}_{32}(\tilde{U}_{2 \rightarrow 13}) = \tilde{X}_{32}(Q, U_{2 \rightarrow 13[Q]}) \triangleq \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}) . \quad (94)$$

This concludes the proof of the converse and thus that of the theorem. \blacksquare

B. Two encoders and three decoders subject to lossless/lossy reconstruction constraints with side information

Consider now the problem described in Fig. 3 where encoder 1 wishes to communicate the *lossless* the source X_1^n to nodes 2 and 3 while encoder 2 wishes to send a *lossy* description of its source X_2^n to nodes 1 and 3 with distortion constraints D_{12} and D_{32} , respectively. In addition to this, the encoders overhead the communication using K communication rounds. This problem can be seen as a generalization of the settings previously investigated in [3], [5].

Theorem 3: The rate-distortion region of the setting described in Fig. 3 is given by the union over all joint probability measures $p_{X_1 X_2 X_3 U_{2 \rightarrow 13} U_{2 \rightarrow 3}}$ satisfying the Markov chain

$$(U_{2 \rightarrow 13}, U_{2 \rightarrow 3}) \text{---} (X_1, X_2) \text{---} X_3 \quad (95)$$

and such that there exists reconstruction mappings:

$$g_{32}(X_1, X_3, U_{2 \rightarrow 13}, U_{2 \rightarrow 3}) = \hat{X}_{32} \quad \text{with} \quad \mathbb{E} \left[d(X_2, \hat{X}_{32}) \right] \leq D_{32}, \quad (96)$$

$$g_{12}(X_1, U_{2 \rightarrow 13}) = \hat{X}_{12} \quad \text{with} \quad \mathbb{E} \left[d(X_2, \hat{X}_{12}) \right] \leq D_{12}, \quad (97)$$

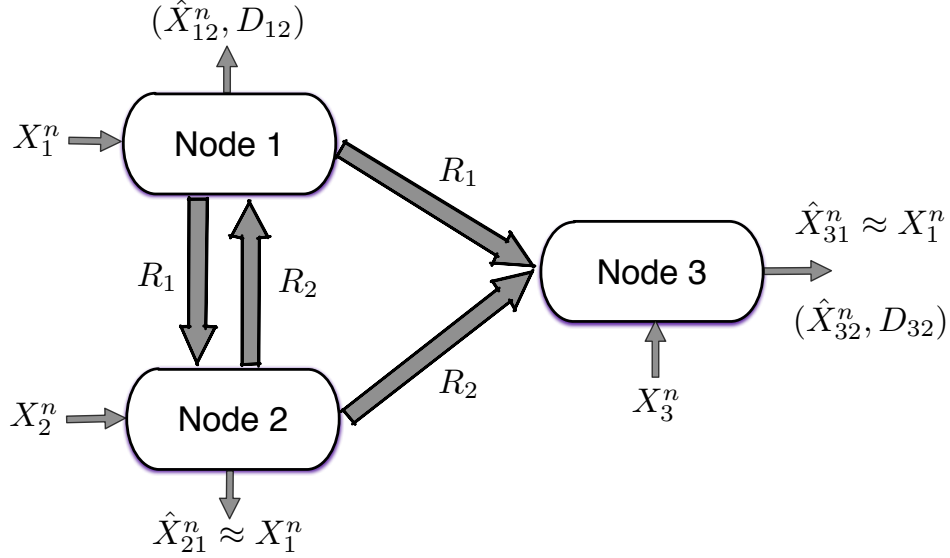


Figure 3: Two encoders and three decoders subject to lossless/lossy reconstruction constraints with side information.

of the set of all tuples satisfying:

$$R_1 \geq H(X_1|X_2) , \quad (98)$$

$$R_2 \geq I(U_{2 \rightarrow 13}; X_2|X_1) + I(U_{2 \rightarrow 3}; X_2|U_{2 \rightarrow 13}X_1X_3) , \quad (99)$$

$$R_1 + R_2 \geq H(X_1|X_3) + I(U_{2 \rightarrow 13}U_{2 \rightarrow 3}; X_2|X_1X_3). \quad (100)$$

The auxiliary random variables have cardinality bounds: $\|\mathcal{U}_{2 \rightarrow 13}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_2\| + 2$, $\|\mathcal{U}_{2 \rightarrow 3}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_2\| \|\mathcal{U}_{2 \rightarrow 13}\| + 1$.

Remark 10: Notice that the rate-distortion region in Theorem 3 is achievable with a single round of interactions $K = 1$, which implies that multiple rounds do not improve the rate-distortion region in this case.

Remark 11: It is worth mentioning that cooperation between encoders reduces the rate needed to communicate the source X_2 while increasing the optimization set of all admissible source descriptions.

Proof: The direct part of the proof follows by choosing:

$$U_{3 \rightarrow 12,l} = U_{1 \rightarrow 3,l} = U_{1 \rightarrow 2,l} = U_{2 \rightarrow 1,l} = U_{3 \rightarrow 1,l} = U_{3 \rightarrow 2,l} = \emptyset, \quad \forall l$$

$$U_{1 \rightarrow 23} \equiv U_{1 \rightarrow 23,1} = X_1, \quad U_{1 \rightarrow 23,l} = U_{2 \rightarrow 13,l} = U_{2 \rightarrow 3,l} = \emptyset \quad \forall l > 1.$$

and $U_{2 \rightarrow 13,1} \equiv U_{2 \rightarrow 13}$ and $U_{2 \rightarrow 3,1} \equiv U_{2 \rightarrow 3}$ are auxiliary random variables that according to Theorem 1 should satisfy:

$$U_{2 \rightarrow 13} \text{---} (X_1, X_2) \text{---} X_3, \quad U_{2 \rightarrow 3} \text{---} (U_{2 \rightarrow 13}, X_1, X_2) \text{---} X_3. \quad (101)$$

Notice, however that these Markov chains are equivalent to (95). From the rate equations in Theorem 1, and the above choices for the auxiliary random variables we obtain:

$$R_{1 \rightarrow 23} > H(X_1 | X_2), \quad (102)$$

$$R_{2 \rightarrow 13} > \max \{I(X_2; U_{2 \rightarrow 13} | X_1), I(X_2; U_{2 \rightarrow 13} | X_1 X_3)\} \quad (103)$$

$$= I(X_2; U_{2 \rightarrow 13} | X_1), \quad (104)$$

$$R_{1 \rightarrow 23} + R_{2 \rightarrow 13} > H(X_1 | X_3) + I(X_2; U_{2 \rightarrow 13} | X_1 X_3), \quad (105)$$

$$R_{2 \rightarrow 3} > I(X_2; U_{2 \rightarrow 3} | U_{2 \rightarrow 13} X_1 X_3). \quad (106)$$

Noticing that $R_1 \equiv R_{1 \rightarrow 23}$ and $R_2 \equiv R_{2 \rightarrow 13} + R_{2 \rightarrow 3}$ the rate-distortion region (34) reduces to the desired region in Theorem 3, where for simplicity we dropped the round index. We now proceed to the proof of the converse.

If a pair of rates (R_1, R_2) and distortions (D_{12}, D_{32}) are admissible for the K -steps interactive cooperative distributed source coding setting described in Fig. 3, then for all $\varepsilon > 0$ there exists $n_0(\varepsilon, K)$, such that $\forall n > n_0(\varepsilon, K)$ there is a K -steps interactive source code $(n, K, \mathcal{F}, \mathcal{G})$ with intermediate rates satisfying:

$$\frac{1}{n} \sum_{l=1}^K \log \|\mathcal{J}_i^l\| \leq R_i + \varepsilon, \quad i \in \{1, 2\} \quad (107)$$

and with average per-letter distortions with respect to the source 2 and perfect reconstruction with respect to the source 1 at all nodes:

$$\mathbb{E} \left[d(X_2^n, \hat{X}_{32}^n) \right] \leq D_{32} + \varepsilon, \quad (108)$$

$$\Pr \left(X_1^n \neq \hat{X}_{21}^n \right) \leq \varepsilon, \quad (109)$$

$$\mathbb{E} \left[d(X_2^n, \hat{X}_{12}^n) \right] \leq D_{12} + \varepsilon, \quad (110)$$

$$\Pr \left(X_1^n \neq \hat{X}_{31}^n \right) \leq \varepsilon, \quad (111)$$

where

$$\hat{X}_{32}^n \equiv g_{32} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right), \quad \hat{X}_{12}^n \equiv g_{12} \left(\mathcal{J}_2^{[1:K]}, X_1^n \right), \quad (112)$$

$$\hat{X}_{31}^n \equiv g_{31} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right), \quad \hat{X}_{21}^n \equiv g_{21} \left(\mathcal{J}_1^{[1:K]}, X_2^n \right). \quad (113)$$

For each $t \in \{1, \dots, n\}$, define random variables $U_{2 \rightarrow 13[t]}$ and $U_{2 \rightarrow 3[t]}$ as follows:

$$U_{2 \rightarrow 13[t]} \triangleq \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_{1[1:t-1]}, X_{1[t+1:n]}, X_{3[1:t-1]} \right), \quad (114)$$

$$U_{2 \rightarrow 3[t]} \triangleq \left(U_{2 \rightarrow 13[t]}, X_{3[t+1:n]}, X_{2[1:t-1]} \right). \quad (115)$$

The fact that these choices of the auxiliary random variables satisfy the Markov chain (95) can be obtained from point 6) in Lemma 10. By the conditions (111) and (109), and Fano's inequality, we have

$$H(X_1^n | \hat{X}_{31}^n) \leq \Pr(X_1^n \neq \hat{X}_{31}^n) \log_2(\|\mathcal{X}_1^n\| - 1) + H_2(\Pr(X_1^n \neq \hat{X}_{31}^n)) \triangleq n\epsilon_n, \quad (116)$$

$$H(X_1^n | \hat{X}_{21}^n) \leq \Pr(X_1^n \neq \hat{X}_{21}^n) \log_2(\|\mathcal{X}_1^n\| - 1) + H_2(\Pr(X_1^n \neq \hat{X}_{21}^n)) \triangleq n\epsilon_n, \quad (117)$$

where $\epsilon_n(\varepsilon) \rightarrow 0$ provided that $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$.

1) *Rate at node 1:* For the first rate, we have

$$n(R_1 + \varepsilon) \geq H(\mathcal{J}_1^{[1:K]}) \quad (118)$$

$$\geq H(\mathcal{J}_1^{[1:K]} | X_2^n) \quad (119)$$

$$\stackrel{(a)}{=} I(\mathcal{J}_1^{[1:K]}; X_1^n | X_2^n) \quad (120)$$

$$= nH(X_1 | X_2) - H(X_1^n | X_2^n \mathcal{J}_1^{[1:K]}) \quad (121)$$

$$\stackrel{(b)}{\geq} nH(X_1 | X_2) - H(X_1^n | \hat{X}_{21}^n) \quad (122)$$

$$\stackrel{(c)}{\geq} n[H(X_1 | X_2) - \epsilon_n], \quad (123)$$

where

- step (a) follows from the fact that by definition of the code the sequence $\mathcal{J}_1^{[1:K]}$ is a function of the both sources (X_1^n, X_2^n) ,
- step (b) follows from the code assumption in (113) that guarantees the existence of a reconstruction function $\hat{X}_{21}^n \equiv g_{21}(\mathcal{J}_1^{[1:K]}, X_2^n)$,
- step (c) follows from Fano's inequality in (117).

2) *Rate at node 2:* For the second rate, we have

$$n(R_2 + \varepsilon) \geq H\left(\mathcal{J}_2^{[1:K]}\right) \quad (124)$$

$$\stackrel{(a)}{=} I\left(\mathcal{J}_2^{[1:K]}; X_1^n X_2^n X_3^n\right) \quad (125)$$

$$\stackrel{(b)}{\geq} I\left(\mathcal{J}_2^{[1:K]}; X_2^n X_3^n | X_1^n\right) \quad (126)$$

$$\stackrel{(c)}{=} I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n X_3^n | X_1^n\right) \quad (127)$$

$$= I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_3^n | X_1^n\right) + I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n\right) \quad (128)$$

$$\stackrel{(d)}{=} \sum_{t=1}^n \left[I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_{3[t]} | X_1^n, X_{3[1:t-1]}\right) \right. \\ \left. + I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_{2[t]} | X_1^n X_3^n X_{2[1:t-1]}\right) \right] \quad (129)$$

$$\stackrel{(e)}{=} \sum_{t=1}^n \left[I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{1[1:t-1]} X_{1[t+1:n]} X_{3[1:t-1]}; X_{3[t]} | X_{1[t]}\right) \right. \\ \left. + I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{1[1:t-1]} X_{1[t+1:n]} X_{3[1:t-1]} X_{3[t+1:n]} X_{2[1:t-1]}; X_{2[t]} | X_{1[t]} X_{3[t]}\right) \right] \quad (130)$$

$$\stackrel{(f)}{=} \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[t]}; X_{3[t]} | X_{1[t]}\right) + I\left(U_{2 \rightarrow 13[t]}; X_{2[t]} | X_{1[t]} X_{3[t]}\right) \right. \\ \left. + I\left(U_{2 \rightarrow 3[t]}; X_{2[t]} | X_{1[t]} X_{3[t]} U_{2 \rightarrow 13[t]}\right) \right] \quad (131)$$

$$= \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[t]}; X_{2[t]} X_{3[t]} | X_{1[t]}\right) + I\left(U_{2 \rightarrow 3[t]}; X_{2[t]} | X_{1[t]} X_{3[t]} U_{2 \rightarrow 13[t]}\right) \right] \quad (132)$$

$$\stackrel{(g)}{=} \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[t]}; X_{2[t]} | X_{1[t]}\right) + I\left(U_{2 \rightarrow 3[t]}; X_{2[t]} | X_{1[t]} X_{3[t]} U_{2 \rightarrow 13[t]}\right) \right] \quad (133)$$

$$\stackrel{(h)}{=} \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[Q]}; X_{2[Q]} | X_{1[Q]}, Q = t\right) \right. \\ \left. + I\left(U_{2 \rightarrow 3[Q]}; X_{2[Q]} | X_{1[Q]} X_{3[Q]} U_{2 \rightarrow 13[Q]}, Q = t\right) \right] \quad (134)$$

$$\stackrel{(i)}{\geq} n \left[I\left(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1\right) + I\left(\tilde{U}_{2 \rightarrow 3}; X_2 | X_1 X_3 \tilde{U}_{2 \rightarrow 13}\right) \right], \quad (135)$$

where

- step (a) follows from the fact that $\mathcal{J}_2^{[1:K]}$ is a function of the sources (X_1^n, X_2^n) ,
- step (b) follows from the non-negativity of mutual information,
- step (c) follows from the fact that $\mathcal{J}_1^{[2:K]}$ is a function of $\mathcal{J}_2^{[1:K]}$ and the source X_1^n ,

- step (d) follows from the chain rule for conditional mutual information,
- step (e) follows from the memoryless property across time of the sources (X_1^n, X_2^n, X_3^n) ,
- step (f) follows from the chain rule for conditional mutual information and the definitions (114) and (115),
- step (g) follows from the Markov chain $U_{2 \rightarrow 13[t]} \ominus (X_{1[t]}, X_{2[t]}) \ominus X_{3[t]}$, for all $t \in \{1, \dots, n\}$,
- step (h) follows from the use of a time sharing random variable Q uniformly distributed over the set $\{1, \dots, n\}$,
- step (i) follows by letting new random variables $\tilde{U}_{2 \rightarrow 13} \triangleq (U_{2 \rightarrow 13[Q]}, Q)$ and $\tilde{U}_{2 \rightarrow 3} \triangleq (U_{2 \rightarrow 3[Q]}, Q)$.

3) *Sum-rate of nodes 1 and 2:* For the sum-rate, we have

$$n(R_1 + R_2 + 2\varepsilon) \geq H(\mathcal{J}_1^{[1:K]}) + H(\mathcal{J}_2^{[1:K]}) \quad (136)$$

$$= H(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}) + I(\mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]}) \quad (137)$$

$$\stackrel{(a)}{=} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n X_3^n X_2^n) + I(\mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]}) \quad (138)$$

$$\stackrel{(b)}{\geq} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n X_2^n | X_3^n) \quad (139)$$

$$= I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n | X_3^n) + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (140)$$

$$= H(X_1^n | X_3^n) - H(X_1^n | \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_3^n) \\ + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (141)$$

$$\stackrel{(c)}{\geq} H(X_1^n | X_3^n) - H(X_1^n | \hat{X}_{31}^n) + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (142)$$

$$\stackrel{(d)}{\geq} n[H(X_1 | X_3) - \epsilon_n] + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (143)$$

$$\stackrel{(e)}{=} \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{1[1:t-1]} X_{1[t+1:n]} X_{3[1:t-1]} X_{3[t+1:n]} X_{2[1:t-1]}; X_{2[t]} | X_{1[t]} X_{3[t]}) \\ + n[H(X_1 | X_3) - \epsilon_n] \quad (144)$$

$$\stackrel{(f)}{=} n[H(X_1 | X_3) - \epsilon_n] + \sum_{t=1}^n I(U_{2 \rightarrow 13[t]} U_{2 \rightarrow 3[t]}; X_{2[t]} | X_{1[t]} X_{3[t]}) \quad (145)$$

$$\stackrel{(g)}{=} n[H(X_1 | X_3) - \epsilon_n + I(U_{2 \rightarrow 13[Q]} U_{2 \rightarrow 3[Q]}; X_{2[Q]} | X_{1[Q]} X_{3[Q]}, Q)] \quad (146)$$

$$\stackrel{(h)}{=} n[H(X_1 | X_3) - \epsilon_n + I(\tilde{U}_{2 \rightarrow 13} \tilde{U}_{2 \rightarrow 3}; X_2 | X_1 X_3)] \quad (147)$$

where

- step (a) follows from the fact that $\mathcal{J}_1^{[1:K]}$ and $\mathcal{J}_2^{[1:K]}$ are functions of the sources (X_1^n, X_2^n, X_3^n) , to emphasize
- step (b) follows non-negativity of mutual information,
- step (c) follows from the code assumption in (113) that guarantees the existence of reconstruction function $\hat{X}_{31}^n \equiv g_{31}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n)$,
- step (d) follows from Fano's inequality in (111),
- step (e) follows from the chain rule of conditional mutual information and the memoryless property across time of the source (X_1^n, X_2^n, X_3^n) ,
- step (f) from follows from the definitions (114) and (115),
- step (g) follows from the use of a time sharing random variable Q uniformly distributed over the set $\{1, \dots, n\}$,
- step (h) follows by letting new random variables $\tilde{U}_{2 \rightarrow 13} \triangleq (U_{2 \rightarrow 13[Q]}, Q)$ and $\tilde{U}_{2 \rightarrow 3} \triangleq (U_{2 \rightarrow 3[Q]}, Q)$.

4) *Distortion at node 1:* Node 1 reconstructs a lossy $\hat{X}_{12}^n \equiv g_{12}(\mathcal{J}_2^{[1:K]}, X_1^n)$. It is clear that we write without loss of generality $\hat{X}_{12}^n \equiv g_{12}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_1^n)$. For each $t \in \{1, \dots, n\}$, define a function $\hat{X}_{12[t]}$ as being the t -th coordinate of this estimate:

$$\hat{X}_{12[t]}(U_{2 \rightarrow 13[t]}, X_{1[t]}) \triangleq g_{12[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_1^n) . \quad (148)$$

The component-wise mean distortion thus verifies

$$D_{12} + \varepsilon \geq \mathbb{E} \left[d \left(X_2, g_{12}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_1^n) \right) \right] \quad (149)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[t]}, \hat{X}_{12[t]}(U_{2 \rightarrow 13[t]}, X_{1[t]}) \right) \right] \quad (150)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{12[Q]}(U_{2 \rightarrow 13[Q]}, X_{1[Q]}) \right) \mid Q = t \right] \quad (151)$$

$$= \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{12[Q]}(U_{2 \rightarrow 13[Q]}, X_{1[Q]}) \right) \right] \quad (152)$$

$$= \mathbb{E} \left[d \left(X_2, \tilde{X}_{12}(\tilde{U}_{2 \rightarrow 13}, X_1) \right) \right] , \quad (153)$$

where we defined function \tilde{X}_{12} by

$$\tilde{X}_{12}(\tilde{U}_{2 \rightarrow 13}, X_1) = \tilde{X}_{12}(Q, U_{2 \rightarrow 13[Q]}, X_{1[Q]}) \triangleq \hat{X}_{12[Q]}(U_{2 \rightarrow 13[Q]}, X_{1[Q]}) . \quad (154)$$

5) *Distortion at node 3*: Node 3 reconstructs a lossy description $\hat{X}_{32}^n \equiv g_{32}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n)$. For each $t \in \{1, \dots, n\}$, define a function $\hat{X}_{32[t]}$ as being the t -th coordinate of this estimate:

$$\hat{X}_{32[t]}(U_{2 \rightarrow 13[t]}, U_{2 \rightarrow 3[t]}, X_{3[t]}) \triangleq g_{32[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n) . \quad (155)$$

The component-wise mean distortion thus verifies

$$D_{32} + \varepsilon \geq \mathbb{E} \left[d \left(X_2, g_{32}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n) \right) \right] \quad (156)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[t]}, \hat{X}_{32[t]}(U_{2 \rightarrow 13[t]}, U_{2 \rightarrow 3[t]}, X_{3[t]}) \right) \right] \quad (157)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}, U_{2 \rightarrow 3[Q]}, X_{3[Q]}) \right) \mid Q = t \right] \quad (158)$$

$$= \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}, U_{2 \rightarrow 3[Q]}, X_{3[Q]}) \right) \right] \quad (159)$$

$$= \mathbb{E} \left[d \left(X_2, \tilde{X}_{32}(\tilde{U}_{2 \rightarrow 13}, \tilde{U}_{2 \rightarrow 3}, X_3) \right) \right] , \quad (160)$$

where we defined function \tilde{X}_{32} by

$$\begin{aligned} \tilde{X}_{32}(\tilde{U}_{2 \rightarrow 13}, \tilde{U}_{2 \rightarrow 3}, X_3) &= \tilde{X}_{32}(Q, U_{2 \rightarrow 13[Q]}, U_{2 \rightarrow 3[Q]}, X_{3[Q]}) \\ &\triangleq \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}, U_{2 \rightarrow 3[Q]}, X_{3[Q]}) . \end{aligned} \quad (161)$$

This concludes the proof of the converse and thus that of the theorem. \blacksquare

C. Two encoders and three decoders subject to lossless/lossy reconstruction constraints, reversal delivery and side information

Consider now the problem described in Fig. 4 where encoder 1 wishes to communicate the *lossless* the source X_1^n to node 2 and a lossy description to node 3. Encoder 2 wishes to send a *lossy* description of its source X_2^n to node 1 and a lossless one to node 3. The corresponding distortion at node 1 and 3 are D_{12} and D_{31} , respectively. In addition to this, the encoders accomplish the communication using K communication rounds. This problem is very similar to the problem described in Fig. 3, with the difference that the decoding at node 3 is inverted.

Theorem 4: The rate-distortion region of the setting described in Fig. 4 is given by the union over all joint probability measures $p_{X_1 X_2 X_3 U_{2 \rightarrow 13}}$ satisfying the Markov chain

$$U_{2 \rightarrow 13} \text{ --- } (X_1, X_2) \text{ --- } X_3 \quad (162)$$

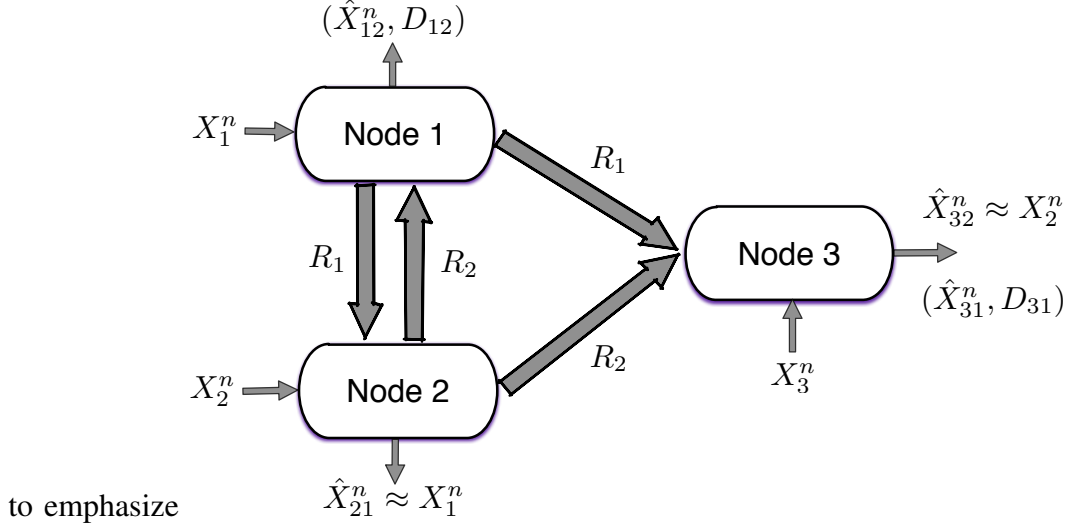


Figure 4: Two encoders and three decoders subject to lossless/lossy reconstruction constraints, reversal delivery and side information.

and such that there exists reconstruction mappings:

$$g_{31}(X_2, X_3, U_{2 \rightarrow 13}) = \hat{X}_{31} \quad \text{with} \quad \mathbb{E} [d(X_1, \hat{X}_{31})] \leq D_{31} , \quad (163)$$

$$g_{12}(X_1, U_{2 \rightarrow 13}) = \hat{X}_{12} \quad \text{with} \quad \mathbb{E} [d(X_2, \hat{X}_{12})] \leq D_{12} , \quad (164)$$

of the set of all tuples satisfying:

$$R_1 \geq H(X_1 | X_2) , \quad (165)$$

$$R_2 \geq I(U_{2 \rightarrow 13}; X_2 | X_1) + H(X_2 | U_{2 \rightarrow 13} X_1 X_3) , \quad (166)$$

$$R_1 + R_2 \geq H(X_1 X_2 | X_3) . \quad (167)$$

The auxiliary random variable has cardinality bounds: $\|\mathcal{U}_{2 \rightarrow 13}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_2\| + 3$.

Remark 12: Notice that the rate-distortion region in Theorem 4 is achievable with a single round of interactions $K = 1$, which implies that multiple rounds do not improve the rate-distortion region in this case.

Remark 13: Notice that, although node 3 requires only the lossy recovery of X_1 , it can in fact recover X_1 perfectly. That is, as node 3 requires the perfect recovery of X_2 , it has the same information that node 2 who recover X_1 perfectly. This explains the sum-rate term. We also see,

that the cooperation helps in the Wyner-Ziv problem that exists between node 2 and 1, with an increasing of the optimization region thanks to the Markov chain (162).

Proof: The direct part of the proof follows by choosing:

$$U_{3 \rightarrow 12, l} = U_{1 \rightarrow 3, l} = U_{1 \rightarrow 2, l} = U_{2 \rightarrow 1, l} = U_{3 \rightarrow 1, l} = U_{3 \rightarrow 2, l} = \emptyset, \quad \forall l$$

$$U_{1 \rightarrow 23} \equiv U_{1 \rightarrow 23, 1} = X_1, \quad U_{2 \rightarrow 3} \equiv U_{2 \rightarrow 3, 1} = X_2, \quad U_{1 \rightarrow 23, l} = U_{2 \rightarrow 13, l} = U_{2 \rightarrow 3, l} = \emptyset \quad \forall l > 1$$

and with $U_{2 \rightarrow 13, 1} \equiv U_{2 \rightarrow 13}$ auxiliary random variable that according to Theorem 1 should satisfy:

$$U_{2 \rightarrow 13} \text{---} (X_1, X_2) \text{---} X_3. \quad (168)$$

From the rate equations in Theorem 1, and the above choices for the auxiliary random variables we obtain:

$$R_{1 \rightarrow 23} > H(X_1 | X_2), \quad (169)$$

$$R_{2 \rightarrow 13} > \max \{I(X_2; U_{2 \rightarrow 13} | X_1), I(X_2; U_{2 \rightarrow 13} | X_1 X_3)\} \quad (170)$$

$$= I(X_2; U_{2 \rightarrow 13} | X_1), \quad (171)$$

$$R_{1 \rightarrow 23} + R_{2 \rightarrow 13} > H(X_1 | X_3) + I(X_2; U_{2 \rightarrow 13} | X_1 X_3), \quad (172)$$

$$R_{2 \rightarrow 3} > H(X_2 | U_{2 \rightarrow 13} X_1 X_3). \quad (173)$$

Noticing that $R_1 \equiv R_{1 \rightarrow 23}$ and $R_2 \equiv R_{2 \rightarrow 13} + R_{2 \rightarrow 3}$ the rate-distortion region (34) reduces to the desired region in Theorem 4, where for simplicity we dropped the round index. We now proceed to the proof of the converse.

If a pair of rates (R_1, R_2) and distortions (D_{12}, D_{31}) are admissible for the K -steps interactive cooperative distributed source coding setting described in Fig. 4, then for all $\varepsilon > 0$ there exists $n_0(\varepsilon, K)$, such that $\forall n > n_0(\varepsilon, K)$ there exists a K -steps interactive source code $(n, K, \mathcal{F}, \mathcal{G})$ with intermediate rates satisfying:

$$\frac{1}{n} \sum_{l=1}^K \log \|\mathcal{J}_i^l\| \leq R_i + \varepsilon, \quad i \in \{1, 2\} \quad (174)$$

and with reconstruction constraints:

$$\mathbb{E} \left[d(X_1^n, \hat{X}_{31}^n) \right] \leq D_{31} + \varepsilon , \quad (175)$$

$$\Pr \left(X_1^n \neq \hat{X}_{21}^n \right) \leq \varepsilon , \quad (176)$$

$$\mathbb{E} \left[d(X_2^n, \hat{X}_{12}^n) \right] \leq D_{12} + \varepsilon , \quad (177)$$

$$\Pr \left(X_2^n \neq \hat{X}_{32}^n \right) \leq \varepsilon , \quad (178)$$

where

$$\hat{X}_{32}^n \equiv g_{32} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right) , \quad \hat{X}_{12}^n \equiv g_{12} \left(\mathcal{J}_2^{[1:K]}, X_1^n \right) , \quad (179)$$

$$\hat{X}_{31}^n \equiv g_{31} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right) , \quad \hat{X}_{21}^n \equiv g_{21} \left(\mathcal{J}_1^{[1:K]}, X_2^n \right) . \quad (180)$$

For each $t \in \{1, \dots, n\}$, define random variables $U_{2 \rightarrow 13[t]}$ as follows:

$$U_{2 \rightarrow 13[t]} \triangleq \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_{1[1:t-1]}, X_{1[t+1:n]}, X_{3[1:t-1]} \right) . \quad (181)$$

Using point 6) in Lemma 10 we can see that this choice satisfies (162). By the conditions (176) and (178), and Fano's inequality, we have

$$H(X_1^n | \hat{X}_{21}^n) \leq \Pr \left(X_1^n \neq \hat{X}_{21}^n \right) \log_2(\|\mathcal{X}_1^n\| - 1) + H_2 \left(\Pr(X_1^n \neq \hat{X}_{21}^n) \right) \triangleq n\epsilon_n , \quad (182)$$

$$H(X_2^n | \hat{X}_{32}^n) \leq \Pr \left(X_2^n \neq \hat{X}_{32}^n \right) \log_2(\|\mathcal{X}_2^n\| - 1) + H_2 \left(\Pr(X_2^n \neq \hat{X}_{32}^n) \right) \triangleq n\epsilon_n , \quad (183)$$

where $\epsilon_n(\varepsilon) \rightarrow 0$ provided that $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$.

1) Rate at node 1: For the first rate, from cut-set arguments similar to the ones used in Theorem 3 and Fano inequality, we can easily obtain:

$$n(R_1 + \varepsilon) \geq n[H(X_1 | X_2) - \epsilon_n] \quad (184)$$

2) *Rate at node 2:* For the second rate, we have

$$n(R_2 + \varepsilon) \geq H\left(\mathcal{J}_2^{[1:K]}\right) \quad (185)$$

$$\stackrel{(a)}{=} I\left(\mathcal{J}_2^{[1:K]}; X_1^n X_2^n X_3^n\right) \quad (186)$$

$$\stackrel{(b)}{\geq} I\left(\mathcal{J}_2^{[1:K]}; X_2^n X_3^n | X_1^n\right) \quad (187)$$

$$\stackrel{(c)}{=} I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n X_3^n | X_1^n\right) \quad (188)$$

$$= I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_3^n | X_1^n\right) + I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n\right) \quad (189)$$

$$\stackrel{(d)}{=} \sum_{t=1}^n \left[I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_{3[t]} | X_1^n, X_{3[1:t-1]}\right) \right. \\ \left. + I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_{2[t]} | X_1^n X_3^n X_{2[1:t-1]}\right) \right] \quad (190)$$

$$\stackrel{(e)}{=} \sum_{t=1}^n \left[I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{1[1:t-1]} X_{1[t+1:n]} X_{3[1:t-1]}; X_{3[t]} | X_{1[t]}\right) \right. \\ \left. + I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{1[1:t-1]} X_{1[t+1:n]} X_{3[1:t-1]} X_{3[t+1:n]} X_{2[1:t-1]}; X_{2[t]} | X_{1[t]} X_{3[t]}\right) \right] \quad (191)$$

$$\stackrel{(f)}{=} \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[t]}; X_{3[t]} | X_{1[t]}\right) + I\left(U_{2 \rightarrow 13[t]} X_{3[t+1:n]} X_{2[1:t-1]}; X_{2[t]} | X_{1[t]} X_{3[t]}\right) \right. \\ \left. = \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[t]}; X_{2[t]} X_{3[t]} | X_{1[t]}\right) \right. \right. \\ \left. \left. + I\left(X_{3[t+1:n]} X_{2[1:t-1]}; X_{2[t]} | X_{1[t]} X_{3[t]} U_{2 \rightarrow 13[t]}\right) \right] \quad (192)$$

$$\stackrel{(g)}{=} \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[t]}; X_{2[t]} | X_{1[t]}\right) + H\left(X_{2[t]} | X_{1[t]} X_{3[t]} U_{2 \rightarrow 13[t]}\right) \right. \\ \left. - H\left(X_{2[t]} | X_{1[t]} X_{3[t:n]} U_{2 \rightarrow 13[t]} X_{2[1:t-1]}\right) \right] \quad (193)$$

$$\stackrel{(h)}{\geq} \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[t]}; X_{2[t]} | X_{1[t]}\right) + H\left(X_{2[t]} | X_{1[t]} X_{3[t]} U_{2 \rightarrow 13[t]}\right) - \epsilon_n \right] \quad (194)$$

$$\stackrel{(i)}{=} \sum_{t=1}^n \left[I\left(U_{2 \rightarrow 13[Q]}; X_{2[Q]} | X_{1[Q]}, Q = t\right) \right. \\ \left. + H\left(X_{2[Q]} | X_{1[Q]} X_{3[Q]} U_{2 \rightarrow 13[Q]}, Q = t\right) - \epsilon_n \right] \quad (195)$$

$$\stackrel{(j)}{\geq} n \left[I\left(\tilde{U}_{2 \rightarrow 13}; X_2 | X_1\right) + H\left(X_2 | X_1 X_3 \tilde{U}_{2 \rightarrow 13}\right) - \epsilon_n \right], \quad (196)$$

where

- step (a) follows from the fact that $\mathcal{J}_2^{[1:K]}$ is a function of the sources (X_1^n, X_2^n) ,
- step (b) follows from the non-negativity of mutual information,
- step (c) follows from the fact that $\mathcal{J}_1^{[2:K]}$ is a function of $\mathcal{J}_2^{[1:K]}$ and the source X_1^n ,
- step (d) follows from the chain rule for conditional mutual information,
- step (e) follows from the memoryless property across time of the sources (X_1^n, X_2^n, X_3^n) ,
- step (f) follows from the definition (181)
- step (g) follows from the Markov chain $U_{2 \rightarrow 13[t]} \ominus (X_{1[t]}, X_{2[t]}) \ominus X_{3[t]}$, for all $t \in \{1, \dots, n\}$ and the usual decomposition of mutual information,
- step (h) follows from the fact that $\Pr(X_{2[t]} \neq \hat{X}_{32[t]}) \leq \epsilon \ \forall t \in \{1, \dots, n\}$, $\hat{X}_{32[t]} \equiv g_{32[t]}(U_{2 \rightarrow 13[t]}, X_{3[t:n]})$ and Fano inequality.
- step (i) follows from the use of a time sharing random variable Q uniformly distributed over the set $\{1, \dots, n\}$,
- step (j) follows by defining new random variables $\tilde{U}_{2 \rightarrow 13} \triangleq (U_{2 \rightarrow 13[Q]}, Q)$.

3) *Sum-rate of nodes 1 and 2:* From cut-set arguments and Fano inequality, we can easily obtain:

$$n(R_1 + R_2 + 2\epsilon) \geq n[H(X_1|X_2X_3) - \epsilon_n] \quad (197)$$

4) *Distortion at node 1:* Node 1 reconstructs a lossy $\hat{X}_{12}^n \equiv g_{12}(\mathcal{J}_2^{[1:K]}, X_1^n)$. For each $t \in \{1, \dots, n\}$, define a function $\hat{X}_{12[t]}$ as the t -th coordinate of this estimate:

$$\hat{X}_{12[t]}(U_{2 \rightarrow 13[t]}, X_{1[t]}) \triangleq g_{12[t]}(\mathcal{J}_2^{[1:K]}, X_1^n) . \quad (198)$$

The component-wise mean distortion thus verifies

$$D_{12} + \epsilon \geq \mathbb{E} \left[d(X_2^n, g_{12}(\mathcal{J}_2^{[1:K]}, X_1^n)) \right] \quad (199)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d(X_{2[t]}, \hat{X}_{12[t]}(U_{2 \rightarrow 13[t]}, X_{1[t]})) \right] \quad (200)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d(X_{2[Q]}, \hat{X}_{12[Q]}(U_{2 \rightarrow 13[Q]}, X_{1[Q]})) \mid Q = t \right] \quad (201)$$

$$= \mathbb{E} \left[d(X_{2[Q]}, \hat{X}_{12[Q]}(U_{2 \rightarrow 13[Q]}, X_{1[Q]})) \right] \quad (202)$$

$$= \mathbb{E} \left[d(X_2, \tilde{X}_{12}(\tilde{U}_{2 \rightarrow 13}, X_1)) \right] , \quad (203)$$

where we defined function \tilde{X}_{12} by

$$\tilde{X}_{12}(\tilde{U}_{2 \rightarrow 13}, X_1) = \tilde{X}_{12}(Q, U_{2 \rightarrow 13[Q]}, X_{1[Q]}) \triangleq \hat{X}_{12[Q]}(U_{2 \rightarrow 13[Q]}, X_{1[Q]}) . \quad (204)$$

5) *Distortion at node 3*: Node 3 reconstructs a lossy description $\hat{X}_{31}^n \equiv g_{31}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n)$. For each $t \in \{1, \dots, n\}$, define a function $\hat{X}_{31[t]}$ as:

$$\bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, \hat{X}_{32[t]}, X_{3[t]}, X_{3[t+1:n]}) \triangleq g_{31[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n) \quad t \in \{1, \dots, n\}. \quad (205)$$

This can be done because $\hat{X}_{32[t]}$ is also a function of $(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n)$. The component-wise mean distortion thus verifies

$$D_{32} + \varepsilon \geq \mathbb{E} \left[d \left(X_1^n, g_{31}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n) \right) \right] \quad (206)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{1[t]}, \bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, \hat{X}_{32[t]}, X_{3[t]}, X_{3[t+1:n]}) \right) \right] \quad (207)$$

$$\stackrel{(a)}{=} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{1[t]}, \bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}, X_{3[t+1:n]}) \right) \right] \quad (208)$$

$$\stackrel{(b)}{\geq} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{1[t]}, \hat{X}_{31[t]}(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}) \right) \right] \quad (209)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{1[Q]}, \hat{X}_{31[Q]}(U_{2 \rightarrow 13[Q]}, X_{2[Q]}, X_{3[Q]}) \right) \mid Q = t \right] \quad (210)$$

$$= \mathbb{E} \left[d \left(X_{1[Q]}, \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}, X_{2[Q]}, X_{3[Q]}) \right) \right] \quad (211)$$

$$= \mathbb{E} \left[d \left(X_1, \tilde{X}_{32}(\tilde{U}_{2 \rightarrow 13}, X_2, X_3) \right) \right], \quad (212)$$

where

- step (a) follows from the fact that $\bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, \hat{X}_{32[t]}, X_{3[t]}, X_{3[t+1:n]})$ can be trivially expressed as a function of $(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}, X_{3[t+1:n]})$ as follows:

$$\bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}, X_{3[t+1:n]}) = \begin{cases} \bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}, X_{3[t+1:n]}) & \text{if } X_{2[t]} = \hat{X}_{32[t]} \\ \bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, \hat{X}_{32[t]}, X_{3[t]}, X_{3[t+1:n]}) & \text{if } X_{2[t]} \neq \hat{X}_{32[t]} \end{cases}$$

- step (b) follows from the fact that $X_{1[t]} \ominus (U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}) \ominus X_{3[t+1:n]}$, which implies that for all $t \in \{1, \dots, n\}$ exists $\hat{X}_{31[t]}(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]})$ such that

$$\mathbb{E} \left[d \left(X_{1[t]}, \hat{X}_{31[t]}(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}) \right) \right] \leq \mathbb{E} \left[d \left(X_{1[t]}, \bar{X}_{31[t]}(U_{2 \rightarrow 13[t]}, X_{2[t]}, X_{3[t]}, X_{3[t+1:n]}) \right) \right]$$

We also defined function \tilde{X}_{32} by

$$\begin{aligned} \tilde{X}_{32}(\tilde{U}_{2 \rightarrow 13}, X_2, X_3) &= \tilde{X}_{32}(Q, U_{2 \rightarrow 13[Q]}, X_{2[Q]}, X_{3[Q]}) \\ &\triangleq \hat{X}_{32[Q]}(U_{2 \rightarrow 13[Q]}, X_{2[Q]}, X_{3[Q]}) \end{aligned} \quad (213)$$

This concludes the proof of the converse and thus that of the theorem. ■

D. Two encoders and three decoders subject to lossy reconstruction constraints with degraded side information

Consider now the problem described in Fig. 5 where encoder 1 has access to X_1 and X_3 and wishes to communicate a *lossy* description of X_1 to nodes 2 and 3 with distortion constraints D_{21} and D_{31} , while encoder 2 wishes to send a *lossy* description of its source X_2^n to nodes 1 and 3 with distortion constraints D_{12} and D_{32} . In addition to this, the encoders overhead the communication using K communication rounds. This problem can be seen as a generalization of the settings previously investigated in [19]. This setup is motivated by the following application. Consider that node 1 transmits a probing signal X_3 which is used to explore a spatial region (i.e. a radar transmitter). After transmission of this probing signal, node 1 measures the response (X_1) at its location. Similarly, in a different location node 2 measures the response X_2 . Responses X_1 and X_2 have to be sent to node 3 (e.g. the fusion center) which has knowledge of the probing signal X_3 and wants to reconstruct a lossy estimate of them. Nodes 1 and 2 cooperate through multiple rounds to accomplish this task.

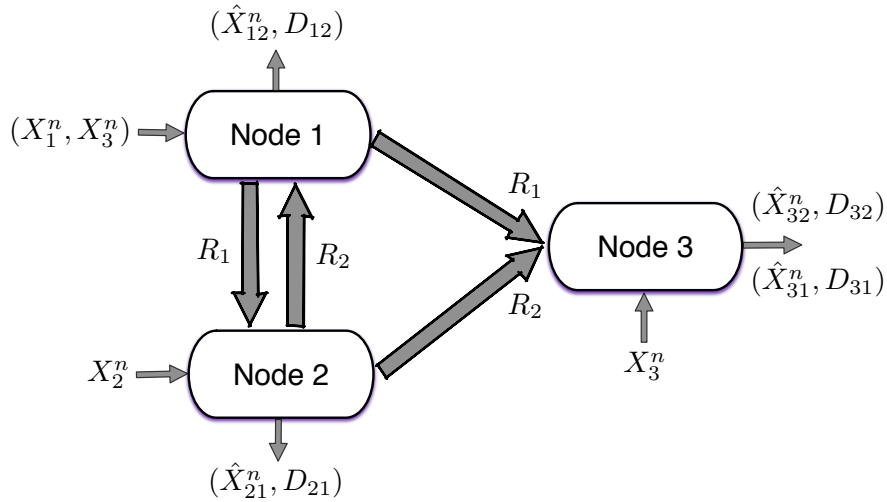


Figure 5: Two encoders and three decoders subject to lossy reconstruction constraints with degraded side information.

Theorem 5: The rate-distortion region of the setting described in Fig. 5 where $X_1 \oplus X_3 \oplus X_2$ form a Markov chain is given by the union over all joint probability measures $p_{X_1 X_2 X_3 \mathcal{W}_{[1, K+1]} U_{1 \rightarrow 3, K}}$

satisfying the following Markov chains:

$$U_{1 \rightarrow 23,l} \text{---} (X_1, X_3, \mathcal{W}_{[1,l]}) \text{---} X_2 , \quad (214)$$

$$U_{2 \rightarrow 13,l} \text{---} (X_2, \mathcal{W}_{[2,l]}) \text{---} (X_1, X_3) , \quad (215)$$

$$U_{1 \rightarrow 3,K} \text{---} (X_1, X_3, \mathcal{W}_{[2,K]}) \text{---} X_2 , \quad (216)$$

for all $l = [1 : K]$, and such that there exist reconstruction mappings:

$$g_{12} (X_1, X_3, U_{1 \rightarrow 3,K}, \mathcal{W}_{[1,K+1]}) = \hat{X}_{12} \quad \text{with} \quad \mathbb{E} [d(X_2, \hat{X}_{12})] \leq D_{12} ,$$

$$g_{21} (X_2, \mathcal{W}_{[1,K+1]}) = \hat{X}_{21} \quad \text{with} \quad \mathbb{E} [d(X_1, \hat{X}_{21})] \leq D_{21} ,$$

$$g_{31} (X_3, \mathcal{W}_{[1,K+1]}, U_{1 \rightarrow 3,K}) = \hat{X}_{31} \quad \text{with} \quad \mathbb{E} [d(X_1, \hat{X}_{31})] \leq D_{31} ,$$

$$g_{32} (X_3, \mathcal{W}_{[1,K+1]}, U_{1 \rightarrow 3,K}) = \hat{X}_{32} \quad \text{with} \quad \mathbb{E} [d(X_2, \hat{X}_{32})] \leq D_{32} ,$$

with $\mathcal{W}_{[1,l]} = \{U_{1 \rightarrow 23,l}, U_{2 \rightarrow 13,l}\}_{k=1}^{l-1}$ for all $l = [1 : K]$,⁹ of the set of all tuples satisfying:

$$R_1 \geq I(\mathcal{W}_{[1,K+1]}; X_1 X_3 | X_2) + I(U_{1 \rightarrow 3,K}; X_1 | \mathcal{W}_{[1,K+1]} X_3) , \quad (217)$$

$$R_2 \geq I(\mathcal{W}_{[1,K+1]}; X_2 | X_3) . \quad (218)$$

The auxiliary random variables have cardinality bounds:

$$\|\mathcal{U}_{1 \rightarrow 23,l}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_3\| \prod_{i=1}^{l-1} \|\mathcal{U}_{1 \rightarrow 23,i}\| \|\mathcal{U}_{2 \rightarrow 13,i}\| + 1, \quad l \in [1 : K] \quad (219)$$

$$\|\mathcal{U}_{2 \rightarrow 13,l}\| \leq \|\mathcal{X}_2\| \|\mathcal{U}_{1 \rightarrow 23,l}\| \prod_{i=1}^{l-1} \|\mathcal{U}_{1 \rightarrow 23,i}\| \|\mathcal{U}_{2 \rightarrow 13,i}\| + 1, \quad l \in [1 : K] \quad (220)$$

$$\|\mathcal{U}_{1 \rightarrow 3,l}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_3\| \prod_{i=1}^K \|\mathcal{U}_{1 \rightarrow 23,i}\| \|\mathcal{U}_{2 \rightarrow 13,i}\| + 3. \quad (221)$$

Remark 14: Notice that multiple rounds are needed to achieve the rate-distortion region in Theorem 5. It is worth to mention that first encoders 1 and 2 cooperate over the K rounds while on the last round only node 1 send a private description to node 3. Because of the Markov chain assumed for the sources we observe the following:

- Only node 1 send a private description to node 3. This is due to the fact that node 3 has better side information than 2.

⁹Notice that $U_{3 \rightarrow 12,l} = \emptyset$ for all l because $R_3 = 0$.

- For the transmissions from node 2, both node 1 and 3 can be thought as an unique node and there is not reason for node 2 to send a private description to node 1 or node 3.
- Notice that there is not sum-rate. Node 3 recovers the descriptions generated at nodes 1 and 2 without resorting to *joint-decoding*. That is, node 3 can recover the descriptions generated at nodes 1 and 2 separately and independently.

Proof: The direct part of the proof follows by choosing:

$$U_{1 \rightarrow 2, l} = U_{2 \rightarrow 1, l} = U_{3 \rightarrow 12, l} = U_{3 \rightarrow 1, l} = U_{3 \rightarrow 2, l} = U_{2 \rightarrow 3, l} = \emptyset, \quad \forall l$$

$$U_{1 \rightarrow 3, l} = \emptyset \quad l < K.$$

and $U_{1 \rightarrow 23, l}$ and $U_{2 \rightarrow 13, l}$ and $U_{1 \rightarrow 3, K}$ are auxiliary random variables that according to Theorem 1 should satisfy the Markov chains (214)-(216). Cumbersome but straightforward calculations allows to obtain the desired results. We now proceed to the proof of the converse.

If a pair of rates (R_1, R_2) and distortions $(D_{12}, D_{21}, D_{31}, D_{32})$ are admissible for the K -steps interactive cooperative distributed source coding setting described in Fig. 5, then for all $\varepsilon > 0$ there exists $n_0(\varepsilon, K)$, such that $\forall n > n_0(\varepsilon, K)$ there exists a K -steps interactive source code $(n, K, \mathcal{F}, \mathcal{G})$ with intermediate rates satisfying:

$$\frac{1}{n} \sum_{l=1}^K \log \|\mathcal{J}_i^l\| \leq R_i + \varepsilon, \quad i \in \{1, 2\} \quad (222)$$

and with average per-letter distortions

$$\mathbb{E} \left[d(X_1^n, \hat{X}_{21}^n) \right] \leq D_{21} + \varepsilon, \quad (223)$$

$$\mathbb{E} \left[d(X_1^n, \hat{X}_{31}^n) \right] \leq D_{31} + \varepsilon, \quad (224)$$

$$\mathbb{E} \left[d(X_2^n, \hat{X}_{12}^n) \right] \leq D_{12} + \varepsilon, \quad (225)$$

$$\mathbb{E} \left[d(X_2^n, \hat{X}_{32}^n) \right] \leq D_{32} + \varepsilon, \quad (226)$$

where

$$\hat{X}_{32}^n \equiv g_{32} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right), \quad \hat{X}_{12}^n \equiv g_{12} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_1^n, X_3^n \right), \quad (227)$$

$$\hat{X}_{31}^n \equiv g_{31} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right), \quad \hat{X}_{21}^n \equiv g_{21} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_2^n \right). \quad (228)$$

For each $t \in \{1, \dots, n\}$, define random variables $(U_{1 \rightarrow 3, [t]}, U_{2 \rightarrow 3, [t]})$ and the sequences of random variables $(U_{1 \rightarrow 23, k, [t]}, U_{2 \rightarrow 13, k, [t]})_{k=[1:K]}$ as follows:

$$U_{1 \rightarrow 23, 1, [t]} \triangleq (\mathcal{J}_1^1, X_{3[1:t-1]}, X_{2[t+1:n]}) \quad , \quad (229)$$

$$U_{2 \rightarrow 13, 1, [t]} \triangleq \mathcal{J}_2^1 \quad , \quad (230)$$

$$U_{1 \rightarrow 23, k, [t]} \triangleq \mathcal{J}_1^k \quad , \quad \forall k = [2 : K] \quad , \quad (231)$$

$$U_{2 \rightarrow 13, k, [t]} \triangleq \mathcal{J}_2^k \quad , \quad \forall k = [2 : K] \quad , \quad (232)$$

$$U_{1 \rightarrow 3, K, [t]} \triangleq X_{3[t+1:n]} \quad , \quad (233)$$

From Corollary 4 in the Appendices we see that these choices satisfy equations (214), (215) and (216).

1) *Rate at node 1:* For the first rate, we have

$$n(R_1 + \varepsilon) \geq H(\mathcal{J}_1^{[1:K]}) \quad (234)$$

$$\stackrel{(a)}{=} I(\mathcal{J}_1^{[1:K]}; X_1^n X_2^n X_3^n) \quad (235)$$

$$\stackrel{(b)}{\geq} I(\mathcal{J}_1^{[1:K]}; X_1^n X_3^n | X_2^n) \quad (236)$$

$$\stackrel{(c)}{=} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n X_3^n | X_2^n) \quad (237)$$

$$= I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n | X_2^n X_3^n) + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_3^n | X_2^n) \quad (238)$$

$$\stackrel{(d)}{=} \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[1:t-1]} X_{2[t+1:n]} X_{3[1:t-1]} X_{3[t+1:n]} X_{1[1:t-1]}; X_{1[t]} | X_{2[t]} X_{3[t]}) \\ + \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{3[1:t-1]} X_{2[1:t-1]} X_{2[t+1:n]}; X_{3[t]} | X_{2[t]}) \quad (239)$$

$$\stackrel{(e)}{\geq} \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t-1]} X_{3[t+1:n]} X_{1[1:t-1]}; X_{1[t]} | X_{2[t]} X_{3[t]}) \\ + \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{3[1:t-1]} X_{2[t+1:n]}; X_{3[t]} | X_{2[t]}) \quad (240)$$

$$\begin{aligned}
&= \sum_{t=1}^n I \left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t-1]}; X_{1[t]} | X_{2[t]} X_{3[t]} \right) \\
&+ \sum_{t=1}^n \left[I \left(X_{3[t+1:n]} X_{1[1:t-1]}; X_{1[t]} | \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t:n]} X_{3[1:t]} \right) \right. \\
&\quad \left. + I \left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{3[1:t-1]} X_{2[t+1:n]}; X_{3[t]} | X_{2[t]} \right) \right] \tag{241}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^n I \left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t-1]}; X_{1[t]} X_{3[t]} | X_{2[t]} \right) \\
&+ \sum_{t=1}^n I \left(X_{3[t+1:n]} X_{1[1:t-1]}; X_{1[t]} | \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t:n]} X_{3[1:t]} \right) \tag{242}
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(f)}{=} \sum_{t=1}^n I \left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t-1]}; X_{1[t]} X_{3[t]} | X_{2[t]} \right) \\
&+ \sum_{t=1}^n I \left(X_{3[t+1:n]} X_{1[1:t-1]}; X_{1[t]} X_{2[t]} | \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t]} \right) \tag{243}
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(g)}{\geq} \sum_{t=1}^n I \left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t-1]}; X_{1[t]} X_{3[t]} | X_{2[t]} \right) \\
&+ \sum_{t=1}^n I \left(X_{3[t+1:n]}; X_{1[t]} | \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t]} \right) \tag{244}
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(h)}{=} \sum_{t=1}^n I \left(U_{1 \rightarrow 23, [1:K], [t]} U_{2 \rightarrow 13, [1:K], [t]}; X_{1[t]} X_{3[t]} | X_{2[t]} \right) \\
&+ \sum_{t=1}^n I \left(U_{1 \rightarrow 3, K, [t]}; X_{1[t]} | U_{1 \rightarrow 23, [1:K], [t]} U_{2 \rightarrow 13, [1:K], [t]} X_{3[t]} \right) \tag{245}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{t=1}^n I \left(U_{1 \rightarrow 23, [1:K], [Q]} U_{2 \rightarrow 13, [1:K], [Q]}; X_{1[Q]} X_{3[Q]} | X_{2[Q]}, Q = t \right) \\
&+ \sum_{t=1}^n I \left(U_{1 \rightarrow 3, K, [Q]}; X_{1[Q]} | U_{1 \rightarrow 23, [1:K], [Q]} U_{2 \rightarrow 13, [1:K], [Q]} X_{3[Q]}, Q = t \right) \tag{246}
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(i)}{=} n \left[I \left(\tilde{U}_{1 \rightarrow 23, [1:K]} \tilde{U}_{2 \rightarrow 13, [1:K]}; X_1 X_3 | X_2 \right) + I \left(\tilde{U}_{1 \rightarrow 3, K}; X_1 | \tilde{U}_{1 \rightarrow 23, [1:K]} \tilde{U}_{2 \rightarrow 13, [1:K]} X_3 \right) \right] \\
&= n \left[I \left(\tilde{\mathcal{W}}_{[1, K+1]}; X_1 X_3 | X_2 \right) + I \left(\tilde{U}_{1 \rightarrow 3, K}; X_1 | \tilde{\mathcal{W}}_{[1, K+1]} X_3 \right) \right] \tag{247}
\end{aligned}$$

where

- step (a) follows from the fact that $\mathcal{J}_1^{[1:K]}$ is a function of the sources (X_1^n, X_2^n, X_3^n) ,
- step (b) follows from the non-negativity of mutual information,
- step (c) follows from the fact that $\mathcal{J}_2^{[1:K]}$ is a function of $\mathcal{J}_1^{[1:K]}$ and the source X_2^n ,

- step (d) follows from the chain rule for conditional mutual information and the memoryless property across time of the sources (X_1^n, X_2^n, X_3^n) ,
- step (e) follows from the non-negativity of mutual information,
- step (f) follows from the Markov chain $X_{2[t]} \ominus (\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{2[t+1:n]} X_{3[1:t]}) \ominus (X_{3[t+1:n]} X_{1[1:t-1]})$ (Corollary 4 in the appendices), for all $t = [1 : n]$ which follows from $X_1 \ominus X_3 \ominus X_2$,
- step (g) follows from the non-negativity of mutual information,
- step (h) follows from definitions (233).
- step (i) follows from the standard time-sharing arguments and the definition of new random variables, (i.e. $\tilde{U}_{1 \rightarrow 23, [1:K]} \triangleq (U_{1 \rightarrow 23, [1:K], [Q]}, Q)$ and $X_1 \triangleq (X_{1[Q]}, Q)$).

and the last step follows from the definition of the past shared common descriptions $\mathcal{W}_{[1,l]} \forall l$. It is also immediate to show that $(\tilde{U}_{1 \rightarrow 23, l}, \tilde{U}_{2 \rightarrow 13, l})$ satisfies the Markov chains in (214)-(216) for all $l \in [1 : K]$.

2) *Rate at node 2:* For the second rate, by following the same steps as before we have

$$n(R_2 + \varepsilon) \geq H(\mathcal{J}_2^{[1:K]}) \quad (248)$$

$$\stackrel{(a)}{=} I(\mathcal{J}_2^{[1:K]}; X_1^n X_2^n X_3^n) \quad (249)$$

$$\stackrel{(b)}{\geq} I(\mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (250)$$

$$\stackrel{(c)}{=} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (251)$$

$$\stackrel{(d)}{=} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_1^n; X_2^n | X_3^n) \quad (252)$$

$$\stackrel{(e)}{\geq} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_2^n | X_3^n) \quad (253)$$

$$\stackrel{(f)}{=} \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{3[1:t-1]} X_{3[t+1:n]} X_{2[t+1:n]}; X_{2[t]} | X_{3[t]}) \quad (254)$$

$$\stackrel{(g)}{\geq} \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_{3[1:t-1]} X_{2[t+1:n]}; X_{2[t]} | X_{3[t]}) \quad (255)$$

$$\stackrel{(h)}{=} \sum_{t=1}^n I(U_{1 \rightarrow 23, [1:K], [t]} U_{2 \rightarrow 13, [1:K], [t]}; X_{2[t]} | X_{3[t]}) \quad (256)$$

$$\stackrel{(i)}{=} nI(\tilde{\mathcal{W}}_{[1, K+1]}; X_2 | X_3) \quad (257)$$

where

- step (a) follows from the fact that $\mathcal{J}_2^{[1:K]}$ is a function of the sources (X_1^n, X_2^n, X_3^n) ,

- step (b) follows from the non-negativity of mutual information,
- step (c) follows from the fact that $\mathcal{J}_1^{[1:K]}$ is a function of $\mathcal{J}_2^{[1:K]}$ and the source (X_1^n, X_3^n) ,
- step (d) follows from the Markov chain $X_1 \ominus X_3 \ominus X_2$.
- step (e) follows from the non-negativity of mutual information,
- step (f) follows from the chain rule for conditional mutual information and the memoryless property across time of the sources (X_1^n, X_2^n, X_3^n) ,
- step (g) follows from the non-negativity of mutual information,
- step (h) follows from definitions (233).
- step (i) follows the definition for $\mathcal{W}_{[1,l]} \forall l$ and from standard time-sharing arguments similar to the ones for rate at node 1.

3) *Distortion at nodes 1 and 2:* Node 1 reconstructs an estimate $\hat{X}_{12}^n \equiv g_{12}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_1^n, X_3^n)$ while node 2 reconstructs $\hat{X}_{21}^n \equiv g_{21}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_2^n)$. For each $t \in \{1, \dots, n\}$, define functions $\hat{X}_{12[t]}$ and $\hat{X}_{21[t]}$ as being the t -th coordinate of the corresponding estimates of \hat{X}_{12}^n and \hat{X}_{21}^n , respectively:

$$\hat{X}_{12[t]}(\mathcal{W}_{[1,K+1],[t]}, U_{1 \rightarrow 3,K,[t]}, X_{1[t]}, X_{3[t]}, X_{1[1:t-1]}, X_{1[t+1:n]}) \triangleq g_{12[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_1^n) \quad (258)$$

$$\hat{X}_{21[t]}(\mathcal{W}_{[1,K+1],[t]}, X_{2[t]}) \triangleq g_{21[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_2^n) \quad (259)$$

The component-wise mean distortions thus verify

$$D_{12} + \varepsilon \geq \mathbb{E} \left[d \left(X_2, g_{12}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_1^n) \right) \right] \quad (260)$$

$$\stackrel{(a)}{=} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[t]}, \hat{X}_{12[t]}(\mathcal{W}_{[1,K+1],[t]}, U_{1 \rightarrow 3,K,[t]}, X_{1[t]}, X_{3[t]}, X_{1[1:t-1]}, X_{1[t+1:n]}) \right) \right] \quad (261)$$

$$\stackrel{(b)}{\geq} \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[t]}, \hat{X}_{12[t]}^*(\mathcal{W}_{[1,K+1],[t]}, U_{1 \rightarrow 3,K,[t]}, X_{1[t]}, X_{3[t]}) \right) \right] \quad (262)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{12[Q]}^*(\mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{1[Q]}, X_{3[Q]}) \right) \mid Q = t \right] \quad (263)$$

$$= \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{12[Q]}^*(\mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{1[Q]}, X_{3[Q]}) \right) \right] \quad (264)$$

$$\stackrel{(c)}{=} \mathbb{E} \left[d \left(X_2, \tilde{X}_{12}(\tilde{\mathcal{W}}_{[1,K+1]}, \tilde{U}_{1 \rightarrow 3,K}, X_1, X_3) \right) \right], \quad (265)$$

where

- step (a) follows from (259),

- step (b) follows from Markov chain $X_{2[t]} \ominus (X_{1[t]}, X_{3[t]}, \mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_{3[1:t-1]}, X_{3[t+1:n]}, X_{2[t+1:n]}) \ominus (X_{1[1:t-1]}, X_{1[t+1:n]}) \forall t = [1 : n]$ (which can be obtained from Corollary 4 in the appendices) and Lemma 9.
- step (c) follows from the following relations:

$$\begin{aligned} \tilde{X}_{12} \left(\tilde{\mathcal{W}}_{[1,K+1]}, \tilde{U}_{1 \rightarrow 3,K}, X_1, X_3 \right) &= \tilde{X}_{12} \left(Q, \mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{1[Q]}, X_{3[Q]} \right) \\ &\triangleq \hat{X}_{12[Q]}^* \left(\mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{1[Q]}, X_{3[Q]} \right) . \end{aligned}$$

By following the very same steps, we can also show that:

$$D_{21} + \varepsilon \geq \mathbb{E} \left[d \left(X_1, g_{21} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_2^n \right) \right) \right] \quad (266)$$

$$= \mathbb{E} \left[d \left(X_1, \tilde{X}_{21} \left(\tilde{\mathcal{W}}_{[1,K+1]}, X_2 \right) \right) \right] , \quad (267)$$

and where we used the Markov chain $X_{2[1:t-1]} \ominus (X_{2[t]}, \mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_{3[1:t-1]}, X_{2[t+1:n]}) \ominus X_{1[t]} \forall t = [1 : n]$ (which can be obtained from Corollary 4 in the appendices) and Lemma 9 and where we define the function \tilde{X}_{21} :

$$\tilde{X}_{21} \left(\tilde{\mathcal{W}}_{[1,K+1]}, X_2 \right) = \tilde{X}_{21} \left(Q, \tilde{\mathcal{W}}_{[1,K+1],[Q]}, X_{2[Q]} \right) \triangleq \hat{X}_{21[Q]}^* \left(\tilde{\mathcal{W}}_{[1,K+1],[Q]}, U_{2 \rightarrow 13[Q]}, X_{2[Q]} \right) .$$

4) *Distortions at node 3:* Node 3 compute lossy reconstructions $\hat{X}_{31}^n \equiv g_{31} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right)$ and $\hat{X}_{32}^n \equiv g_{32} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right)$. For each $t \in \{1, \dots, n\}$, define functions $\hat{X}_{31[t]}$ and $\hat{X}_{32[t]}$ as being the t -th coordinate of the corresponding estimates of \hat{X}_{31}^n and \hat{X}_{32}^n , respectively:

$$\hat{X}_{31[t]} \left(\mathcal{W}_{[1,K+1],[t]}, U_{1 \rightarrow 3,K,[t]}, X_{3[t]} \right) \triangleq g_{31[t]} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right) , \quad (268)$$

$$\hat{X}_{32[t]} \left(\mathcal{W}_{[1,K+1],[t]}, U_{1 \rightarrow 3,K,[t]}, X_{3[t]} \right) \triangleq g_{32[t]} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right) . \quad (269)$$

The component-wise mean distortions thus verify

$$D_{31} + \varepsilon \geq \mathbb{E} \left[d \left(X_1, g_{31} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n \right) \right) \right] \quad (270)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{1[t]}, \hat{X}_{31[t]} \left(\mathcal{W}_{[1,K+1],[t]}, U_{1 \rightarrow 3,K,[t]}, X_{3[t]} \right) \right) \right] \quad (271)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{1[Q]}, \hat{X}_{31[Q]} \left(\mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{3[Q]} \right) \mid Q = t \right) \right] \quad (272)$$

$$= \mathbb{E} \left[d \left(X_{1[Q]}, \hat{X}_{31[Q]} \left(\mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{3[Q]} \right) \right) \right] \quad (273)$$

$$= \mathbb{E} \left[d \left(X_1, \tilde{X}_{31} \left(\tilde{\mathcal{W}}_{[1,K+1]}, \tilde{U}_{1 \rightarrow 3,K}, X_3 \right) \right) \right] , \quad (274)$$

where the last step follows by defining the function \tilde{X}_{31} by

$$\begin{aligned}\tilde{X}_{31}(\tilde{\mathcal{W}}_{[1,K+1]}, \tilde{U}_{1 \rightarrow 3,K}, X_3) &= \tilde{X}_{31}(Q, \mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{3[Q]}) \\ &\triangleq \hat{X}_{31[Q]}(\mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{3[Q]}) .\end{aligned}$$

By following the very same steps, we can also show that:

$$D_{32} + \varepsilon \geq \mathbb{E} \left[d \left(X_2, g_{32}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, X_3^n) \right) \right] \quad (275)$$

$$= \mathbb{E} \left[d \left(X_2, \tilde{X}_{32}(\tilde{\mathcal{W}}_{[1,K+1]}, \tilde{U}_{1 \rightarrow 3,K}, X_3) \right) \right] , \quad (276)$$

and where we define the function \tilde{X}_{32} by

$$\begin{aligned}\tilde{X}_{32}(\tilde{\mathcal{W}}_{[1,K+1]}, \tilde{U}_{1 \rightarrow 3,K}, X_3) &= \tilde{X}_{32}(Q, \mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{3[Q]}) \\ &\triangleq \hat{X}_{32[Q]}(\mathcal{W}_{[1,K+1],[Q]}, U_{1 \rightarrow 3,K,[Q]}, X_{3[Q]}) .\end{aligned}$$

Cooperative distributed source coding with two distortion criteria under reversal delivery and side information This concludes the proof of the converse and thus that of the theorem. ■

E. Three encoders and three decoders subject to lossless/lossy reconstruction constraints with degraded side information

Consider now the problem described in Fig. 6 where encoder 1 wishes to communicate the *lossless* the source X_1^n to nodes 2 and 3 while encoder 2 wishes to send a *lossy* description of its source X_2^n to node 3 with distortion constraints D_{32} and encoder 3 wishes to send a *lossy* description of its source X_3^n to node 2 with distortion constraints D_{23} . In addition to this, the encoders perform the communication using K communication rounds. This problem can be seen as a generalization of the settings previously investigated in [10]. This setting can model a problem in which node 1 generate a process X_1 . This process physically propagates to the locations where nodes 2 and 3 are. These nodes measure X_2 and X_3 respectively. If node 2 is closer to node 1 than node 3, we can assume that $X_1 \ominus X_2 \ominus X_3$. Nodes 2 and 3 then interact between them and with node 1, in order to reconstruct X_1 in lossless fashion and X_2 and X_3 with some distortion level.

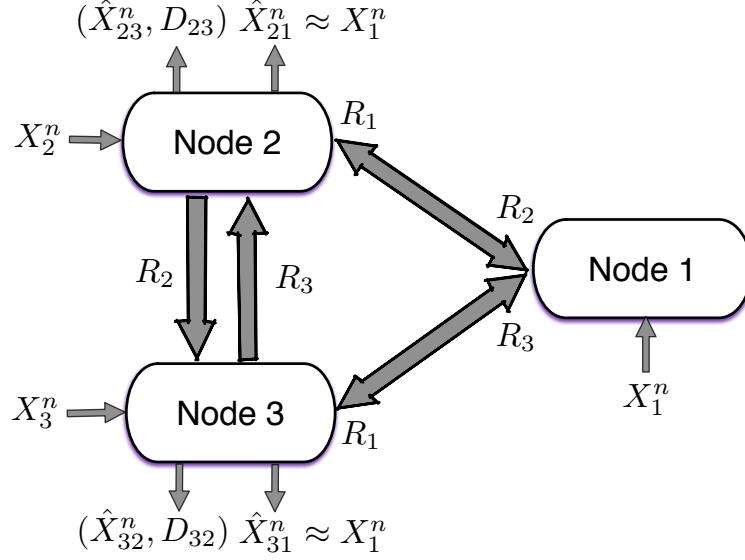


Figure 6: Three encoders and three decoders subject to lossless/lossy reconstruction constraints with degraded side information.

Theorem 6: The rate-distortion region of the setting described in Fig. 6 where $X_1 \ominus X_2 \ominus X_3$ form a Markov chain is given by the union over all joint probability measures $p_{X_1 X_2 X_3 U_{3 \rightarrow 2, [1:K]} U_{2 \rightarrow 3, [1:K]}}$ satisfying the Markov chains

$$U_{2 \rightarrow 3, l} \ominus (X_1, X_2, \mathcal{V}_{[23, l, 2]}) \ominus X_3, \quad (277)$$

$$U_{3 \rightarrow 2, l} \ominus (X_1, X_3, \mathcal{V}_{[23, l, 3]}) \ominus X_2, \quad (278)$$

$\forall l \in [1 : K]$, and such that there exist reconstruction mappings:

$$g_{23}(X_1, X_2, \mathcal{V}_{[23, K+1, 2]}) = \hat{X}_{23} \quad \text{with} \quad \mathbb{E} [d(X_3, \hat{X}_{23})] \leq D_{23},$$

$$g_{32}(X_1, X_3, \mathcal{V}_{[23, K+1, 2]}) = \hat{X}_{32} \quad \text{with} \quad \mathbb{E} [d(X_2, \hat{X}_{32})] \leq D_{32},$$

of the set of all tuples satisfying:

$$R_1 \geq H(X_1 | X_2), \quad (279)$$

$$R_2 \geq I(\mathcal{V}_{[23, K+1, 2]}; X_2 | X_1 X_3), \quad (280)$$

$$R_3 \geq I(\mathcal{V}_{[23, K+1, 2]}; X_3 | X_1 X_2), \quad (281)$$

$$R_1 + R_2 \geq H(X_1 | X_3) + I(\mathcal{V}_{[23, K+1, 2]}; X_2 | X_1 X_3), \quad (282)$$

The auxiliary random variables have cardinality bounds:

$$\|\mathcal{U}_{2 \rightarrow 3, l}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_2\| \prod_{i=1}^{l-1} \|\mathcal{U}_{2 \rightarrow 3, i}\| \|\mathcal{U}_{3 \rightarrow 2, i}\| + 1, \quad l \in [1 : K] \quad (283)$$

$$\|\mathcal{U}_{3 \rightarrow 2, l}\| \leq \|\mathcal{X}_1\| \|\mathcal{X}_3\| \|\mathcal{U}_{2 \rightarrow 3, l}\| \prod_{i=1}^{l-1} \|\mathcal{U}_{2 \rightarrow 3, i}\| \|\mathcal{U}_{3 \rightarrow 2, i}\| + 1, \quad l \in [1 : K] \quad (284)$$

Remark 15: Theorem 6 shows that several exchanges between nodes 2 and 3 can be helpful. Node 1 transmit only once at the beginning its full source.

Proof: The direct part of the proof follows according to Theorem 1 by choosing:

$$U_{1 \rightarrow 3, l} = U_{1 \rightarrow 2, l} = U_{3 \rightarrow 1, l} = U_{2 \rightarrow 1, l} = U_{3 \rightarrow 12, l} = \emptyset \quad \forall l \in [1 : K]:$$

$$U_{1 \rightarrow 23, l} = U_{2 \rightarrow 13, l} = \emptyset, \quad \forall l \in [2 : K]$$

and $U_{1 \rightarrow 23, 1} = U_{2 \rightarrow 13, 1} = X_1$. The remanding auxiliary random variables satisfy $\forall l \in [1 : K]$:

$$U_{2 \rightarrow 3, l} \ominus (X_1, X_2, \mathcal{V}_{[23, l, 2]}) \ominus X_3, \quad (285)$$

$$U_{3 \rightarrow 2, l} \ominus (X_1, X_3, \mathcal{V}_{[23, l, 3]}) \ominus X_2. \quad (286)$$

If a pair of tuple (R_1, R_2, R_3) and distortions (D_{23}, D_{32}) are admissible for the K -steps interactive cooperative distributed source coding setting described in Fig. 6, then for all $\varepsilon > 0$ there exists $n_0(\varepsilon, K)$, such that $\forall n > n_0(\varepsilon, K)$ there exists a K -steps interactive source code $(n, K, \mathcal{F}, \mathcal{G})$ with intermediate rates satisfying:

$$\frac{1}{n} \sum_{l=1}^K \log \|\mathcal{J}_i^l\| \leq R_i + \varepsilon, \quad i \in \{1, 2, 3\} \quad (287)$$

and with average per-letter distortions with respect to the source 2 and perfect reconstruction with respect to the source 1 at all nodes:

$$\mathbb{E} \left[d(X_2^n, \hat{X}_{32}^n) \right] \leq D_{32} + \varepsilon, \quad (288)$$

$$\Pr \left(X_1^n \neq \hat{X}_{21}^n \right) \leq \varepsilon, \quad (289)$$

$$\mathbb{E} \left[d(X_3^n, \hat{X}_{23}^n) \right] \leq D_{23} + \varepsilon, \quad (290)$$

$$\Pr \left(X_1^n \neq \hat{X}_{31}^n \right) \leq \varepsilon, \quad (291)$$

where

$$\hat{X}_{32}^n \equiv g_{32} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_3^n \right), \quad \hat{X}_{23}^n \equiv g_{23} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_2^n \right), \quad (292)$$

$$\hat{X}_{31}^n \equiv g_{31} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_3^n \right), \quad \hat{X}_{21}^n \equiv g_{21} \left(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_2^n \right). \quad (293)$$

For each $t \in \{1, \dots, n\}$ and $l \in [1 : K]$, we define random variables $U_{2 \rightarrow 3, l, [t]}$ and $U_{3 \rightarrow 2, l, [t]}$ as follows:

$$U_{2 \rightarrow 3, 1, [t]} \triangleq \left(\mathcal{J}_1^1, \mathcal{J}_2^1, X_{1[1:t-1]}, X_{1[t+1:n]}, X_{2[t+1:n]}, X_{3[1:t-1]} \right), \quad (294)$$

$$U_{2 \rightarrow 3, l, [t]} \triangleq \left(\mathcal{J}_1^l, \mathcal{J}_2^l \right), \quad l \in [2 : K] \quad (295)$$

$$U_{3 \rightarrow 2, l, [t]} \triangleq \mathcal{J}_3^l, \quad l \in [1 : K]. \quad (296)$$

These auxiliary random variables satisfy the Markov conditions (285) and (286), which can be verified from Lemma 11 in the appendices. By the conditions (291) and (289), and Fano's inequality [21], we have

$$H(X_1^n | \hat{X}_{31}^n) \leq \Pr(X_1^n \neq \hat{X}_{31}^n) \log_2(\|\mathcal{X}_1^n\| - 1) + H_2\left(\Pr(X_1^n \neq \hat{X}_{31}^n)\right) \triangleq n\epsilon_n, \quad (297)$$

$$H(X_1^n | \hat{X}_{21}^n) \leq \Pr(X_1^n \neq \hat{X}_{21}^n) \log_2(\|\mathcal{X}_1^n\| - 1) + H_2\left(\Pr(X_1^n \neq \hat{X}_{21}^n)\right) \triangleq n\epsilon_n, \quad (298)$$

where $\epsilon_n(\varepsilon) \rightarrow 0$ provided that $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$.

1) *Rate at node 1:* For the first rate, we have

$$n(R_1 + \varepsilon) \geq H\left(\mathcal{J}_1^{[1:K]}\right) \quad (299)$$

$$\geq H\left(\mathcal{J}_1^{[1:K]} | X_2^n X_3^n\right) \quad (300)$$

$$\stackrel{(a)}{=} H\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]} | X_2^n X_3^n\right) \quad (301)$$

$$\stackrel{(b)}{=} I\left(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}; X_1^n | X_2^n X_3^n\right) \quad (302)$$

$$\geq nH(X_1 | X_2 X_3) - H\left(X_1^n | X_2^n \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}\right) \quad (303)$$

$$\stackrel{(c)}{\geq} nH(X_1 | X_2 X_3) - H(X_1^n | \hat{X}_{21}^n) \quad (304)$$

$$\stackrel{(d)}{\geq} n[H(X_1 | X_2 X_3) - \epsilon_n], \quad (305)$$

$$\stackrel{(e)}{=} n[H(X_1 | X_2) - \epsilon_n], \quad (306)$$

where

- step (a) follows from the fact that by definition of the code the sequence $\mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}$ are functions of $(\mathcal{J}_1^{[1:K]}, X_2^n, X_3^n)$,
- step (b) follows from the fact that by definition of the code the sequences $(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]})$ are functions of the sources (X_1^n, X_2^n, X_3^n) ,
- step (c) follows from the code assumption in (293) that guarantees the existence of a reconstruction function $\hat{X}_{21}^n \equiv g_{21}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_2^n)$,
- step (d) follows from Fano's inequality in (298),
- step (e) follows from the assumption that $X_1 \text{---} X_2 \text{---} X_3$ form a Markov chain.

2) *Rate at nodes 2 and 3:* For the second rate, we have

$$n(R_2 + \varepsilon) \geq H(\mathcal{J}_2^{[1:K]}) \quad (307)$$

$$\stackrel{(a)}{=} I(\mathcal{J}_2^{[1:K]}; X_1^n X_2^n X_3^n) \quad (308)$$

$$\stackrel{(b)}{\geq} I(\mathcal{J}_2^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (309)$$

$$\stackrel{(c)}{=} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (310)$$

$$\stackrel{(d)}{=} \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}; X_{2[t]} | X_1^n X_3^n X_{2[t+1:n]}) \quad (311)$$

$$\stackrel{(e)}{\geq} \sum_{t=1}^n I(\mathcal{V}_{[23, K+1, 2][t]}; X_{2[t]} | X_{1[t]} X_{3[t]}) \quad (312)$$

$$\stackrel{(f)}{=} \sum_{t=1}^n I(\mathcal{V}_{[23, K+1, 2][Q]}; X_{2[Q]} | X_{1[Q]} X_{3[Q]}, Q = t) \quad (313)$$

$$\stackrel{(g)}{\geq} nI(\tilde{\mathcal{V}}_{[23, K+1, 2]}; X_2 | X_1 X_3) \quad (314)$$

where

- step (a) follows from the fact that $\mathcal{J}_2^{[1:K]}$ is a function of the sources (X_1^n, X_2^n, X_3^n) ,
- step (b) follows from the non-negativity of mutual information,
- step (c) follows from the fact that $(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]})$ are functions of the sources (X_1^n, X_2^n, X_3^n) ,
- step (d) follows from the chain rule for conditional mutual information,
- step (e) follows from the definitions (294), (295) and (296), the memoryless property of the sources and the non-negativity of mutual information,

- step (f) follows from the use of a time sharing random variable Q uniformly distributed over the set $\{1, \dots, n\}$,
- step (g) follows by letting a new random variable $\tilde{\mathcal{V}}_{[23, K+1, 2]} \triangleq (\mathcal{V}_{[23, K+1, 2][Q]}, Q)$.

By following similar steps, it is not difficult to check that

$$n(R_3 + \varepsilon) \geq \sum_{t=1}^n I(\mathcal{V}_{[23, K+1, 2][t]}; X_{3[t]} | X_{1[t]} X_{2[t]}) \quad (315)$$

$$= \sum_{t=1}^n I(\mathcal{V}_{[23, K+1, 2][Q]}; X_{3[Q]} | X_{1[Q]} X_{2[Q]}, Q = t) \quad (316)$$

$$\geq nI(\tilde{\mathcal{V}}_{[23, K+1, 2]}; X_3 | X_1 X_2) . \quad (317)$$

3) *Sum-rate of nodes 1 and 2:* For the sum-rate, we have

$$n(R_1 + R_2 + 2\varepsilon) \geq H(\mathcal{J}_1^{[1:K]}) + H(\mathcal{J}_2^{[1:K]}) \quad (318)$$

$$\geq H(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}) \quad (319)$$

$$\stackrel{(a)}{=} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n X_3^n X_2^n) \quad (320)$$

$$\stackrel{(b)}{\geq} I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n X_2^n | X_3^n) \quad (321)$$

$$= I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]}; X_1^n | X_3^n) + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (322)$$

$$= H(X_1^n | X_3^n) - H(X_1^n | \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} X_3^n) \\ + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (323)$$

$$\stackrel{(c)}{\geq} H(X_1^n | X_3^n) - H(X_1^n | \hat{X}_{31}^n) + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (324)$$

$$\stackrel{(d)}{\geq} n[H(X_1 | X_3) - \epsilon_n] + I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]}; X_2^n | X_1^n X_3^n) \quad (325)$$

$$\stackrel{(e)}{\geq} \sum_{t=1}^n I(\mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \mathcal{J}_3^{[1:K]} X_{1[1:t-1]} X_{1[t+1:n]} X_{3[1:t-1]} X_{2[t+1:n]}; X_{2[t]} | X_{1[t]} X_{3[t]}) \\ + n[H(X_1 | X_3) - \epsilon_n] \quad (326)$$

$$\stackrel{(f)}{=} n[H(X_1 | X_3) - \epsilon_n] + \sum_{t=1}^n I(\mathcal{V}_{[23, K+1, 2][t]}; X_{2[t]} | X_{1[t]} X_{3[t]}) \quad (327)$$

$$\stackrel{(g)}{=} n[H(X_1 | X_3) - \epsilon_n + I(\mathcal{V}_{[23, K+1, 2][Q]}; X_{2[Q]} | X_{1[Q]} X_{3[Q]}, Q)] \quad (328)$$

$$\stackrel{(h)}{=} n[H(X_1 | X_3) - \epsilon_n + I(\tilde{\mathcal{V}}_{[23, K+1, 2]}; X_2 | X_1 X_3)] , \quad (329)$$

where

- step (a) follows from the fact that $\mathcal{J}_1^{[1:K]}$ and $\mathcal{J}_2^{[1:K]}$ are functions of the sources (X_1^n, X_2^n, X_3^n) ,
- step (b) follows non-negativity of mutual information,
- step (c) follows from the code assumption in (293) that guarantees the existence of reconstruction function $\hat{X}_{31}^n \equiv g_{31}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_3^n)$,
- step (d) follows from Fano's inequality in (291),
- step (e) follows from the chain rule of conditional mutual information and the memoryless property across time of the sources (X_1^n, X_2^n, X_3^n) , and non-negativity of mutual information,
- step (f) from follows from the definitions (294) and (295),
- step (g) follows from the use of a time sharing random variable Q uniformly distributed over the set $\{1, \dots, n\}$,
- step (h) follows from $\tilde{\mathcal{V}}_{[23, K+1, 2]} \triangleq (\tilde{\mathcal{V}}_{[23, K+1, 2][Q]}, Q)$.

4) *Distortion at node 2:* Node 2 reconstructs a lossy $\hat{X}_{23}^n \equiv g_{23}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_2^n)$.

For each $t \in \{1, \dots, n\}$, define a function $\hat{X}_{23[t]}$ as being the t -th coordinate of this estimate:

$$\hat{X}_{23[t]}(\mathcal{V}_{[23, K+1, 2][t]}, X_{2[t]}) \triangleq g_{23[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_2^n) . \quad (330)$$

The component-wise mean distortion thus verifies

$$D_{23} + \varepsilon \geq \mathbb{E} \left[d \left(X_3, g_{23}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_2^n) \right) \right] \quad (331)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{3[t]}, \hat{X}_{23[t]}(\mathcal{V}_{[23, K+1, 2][t]}, X_{2[t]}) \right) \right] \quad (332)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{3[Q]}, \hat{X}_{23[Q]}(\mathcal{V}_{[23, K+1, 2][Q]}, X_{2[Q]}) \right) \mid Q = t \right] \quad (333)$$

$$= \mathbb{E} \left[d \left(X_{3[Q]}, \hat{X}_{23[Q]}(\mathcal{V}_{[23, K+1, 2][Q]}, X_{2[Q]}) \right) \right] \quad (334)$$

$$= \mathbb{E} \left[d \left(X_3, \tilde{X}_{23}(\tilde{\mathcal{V}}_{[23, K+1, 2]}, X_2) \right) \right] , \quad (335)$$

where we defined function \tilde{X}_{23} by

$$\tilde{X}_{23}(\tilde{\mathcal{V}}_{[23, K+1, 2]}, X_2) = \tilde{X}_{23}(Q, \mathcal{V}_{[23, K+1, 2][Q]}, X_{2[Q]}) \triangleq \hat{X}_{23[Q]}(\mathcal{V}_{[23, K+1, 2][Q]}, X_{2[Q]}) . \quad (336)$$

5) *Distortion at node 3:* Node 3 reconstructs a lossy description $\hat{X}_{32}^n \equiv g_{32}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_3^n)$. For each $t \in \{1, \dots, n\}$, define a function $\hat{X}_{32[t]}$ as being the t -th coordinate of this estimate:

$$\hat{X}_{32[t]}(\mathcal{V}_{[23, K+1, 2][t]}, X_{3[t]}) \triangleq g_{32[t]}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_3^n) . \quad (337)$$

The component-wise mean distortion thus verifies

$$D_{32} + \varepsilon \geq \mathbb{E} \left[d \left(X_2, g_{32}(\mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]}, \mathcal{J}_3^{[1:K]}, X_3^n) \right) \right] \quad (338)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[t]}, \hat{X}_{32[t]}(\mathcal{V}_{[23, K+1, 2][t]}, X_{3[t]}) \right) \right] \quad (339)$$

$$= \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{32[Q]}(\mathcal{V}_{[23, K+1, 2][Q]}, X_{3[Q]}) \right) \mid Q = t \right] \quad (340)$$

$$= \mathbb{E} \left[d \left(X_{2[Q]}, \hat{X}_{32[Q]}(\mathcal{V}_{[23, K+1, 2][Q]}, X_{3[Q]}) \right) \right] \quad (341)$$

$$= \mathbb{E} \left[d \left(X_2, \tilde{X}_{32}(\tilde{\mathcal{V}}_{[23, K+1, 2]}, X_3) \right) \right], \quad (342)$$

where we defined function \tilde{X}_{32} by

$$\begin{aligned} \tilde{X}_{32}(\tilde{\mathcal{V}}_{[23, K+1, 3]}, X_3) &= \tilde{X}_{32}(Q, \mathcal{V}_{[23, K+1, 3][Q]}, X_{3[Q]}) \\ &\triangleq \hat{X}_{32[Q]}(\mathcal{V}_{[23, K+1, 3][Q]}, X_{3[Q]}) . \end{aligned} \quad (343)$$

This concludes the proof of the converse and thus that of the theorem. ■

VI. DISCUSSION

A. Numerical example

In order to obtain further insight into the gains obtained from cooperation, we consider the case of two encoders and one decoder subject to lossy/lossless reconstruction constraints without side information in which the sources are distributed according to:

$$\begin{aligned} p_{X_1 X_2}(x_1, x_2) &= \alpha \mathbb{1}\{x_1 = 1\} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x_2^2}{2\sigma_1^2}\right) \\ &\quad + (1 - \alpha) \mathbb{1}\{x_1 = 0\} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x_2^2}{2\sigma_0^2}\right). \end{aligned} \quad (344)$$

This model yields a mixed between discrete and continuous components. We observe that X_1 follows a *Bernoulli* distribution with parameter $\alpha \in [0 : 1]$ while X_2 given X_1 follows a Gaussian distribution with different variance according to the value of $X_1 \in \{0, 1\}$. In this sense, X_2 follows a Gaussian mixture distribution¹⁰.

¹⁰Although the inner bound region in Theorem 1 is strictly valid for discrete sources with finite alphabets, the Gaussian distribution is sufficiently well-behaved to apply a uniform quantization procedure prior to the application of the results of Theorem 1. Then, a limiting argument using a sequence of decreasing quantization step-sizes will deliver the desired result. See chapter 3 in [22].

The optimal rate-distortion region for this case was characterized in Theorem 2 and can be alternatively written as:

$$\begin{aligned} \mathcal{R}(D) = \bigcup_{p \in \mathcal{L}} \Big\{ (R_1, R_2) : & R_1 > H(X_1|X_2) , \\ & R_2 > I(X_2; U|X_1) , \\ & R_1 + R_2 > H(X_1) + I(X_2; U|X_1) \Big\} , \end{aligned} \quad (345)$$

where

$$\mathcal{L} = \{p_{U|X_1X_2} : \text{there exists } (x_1, u) \mapsto g(x_1, u) \text{ such that } \mathbb{E}[d(X_2, g(X_1, U))] \leq D\} . \quad (346)$$

The corresponding non-cooperative region for the same problem was characterized in [5]:

$$\mathcal{R}(D) = \bigcup_{p \in \mathcal{L}^*} \Big\{ (R_1, R_2) : R_1 > H(X_1|U) , \quad (347)$$

$$R_2 > I(X_2; U|X_1) , \quad (348)$$

$$R_1 + R_2 > H(X_1) + I(X_2; U|X_1) \Big\} , \quad (349)$$

where

$$\mathcal{L}^* = \{p_{U|X_2} : \text{there exists } (x_1, u) \mapsto g(x_1, u) \text{ such that } \mathbb{E}[d(X_2, g(X_1, U))] \leq D\} . \quad (350)$$

From the previous expressions, it is evident that the cooperative case offers some gains with respect to the non-cooperative setup. This is clearly evidenced from the lower limit in R_1 and the fact that $\mathcal{L}^* \subseteq \mathcal{L}$. We have the following result.

Theorem 7 (Cooperative region for mixed discrete/continuous source): Assume the source distribution is given by (344) and that, without loss of generality, $\sigma_0^2 \leq \sigma_1^2$. The rate-distortion region

from Theorem 2 can be written as:

$$\begin{aligned}
R_1 &> \frac{1-\alpha}{\sqrt{2\pi}\sigma_0} \int_{-\infty}^{\infty} \exp\left(-\frac{x_2^2}{2\sigma_0^2}\right) H_2(g(x_2)) dx_2 \\
&\quad + \frac{\alpha}{\sqrt{2\pi}\sigma_1} \int_{-\infty}^{\infty} \exp\left(-\frac{x_2^2}{2\sigma_1^2}\right) H_2(g(x_2)) dx_2, \\
R_2 &> \begin{cases} \frac{1}{2} \log\left(\frac{\sigma_0^{2(1-\alpha)} \sigma_1^{2\alpha}}{D}\right) & D \leq \sigma_0^2 \\ \frac{\alpha}{2} \left[\log\left(\frac{\alpha \sigma_1^2}{D - (1-\alpha)\sigma_0^2}\right) \right]^+ & D > \sigma_0^2 \end{cases}, \\
R_1 + R_2 &> \begin{cases} H_2(\alpha) + \frac{1}{2} \log\left(\frac{\sigma_0^{2(1-\alpha)} \sigma_1^{2\alpha}}{D}\right) & D \leq \sigma_0^2 \\ H_2(\alpha) + \frac{\alpha}{2} \left[\log\left(\frac{\alpha \sigma_1^2}{D - (1-\alpha)\sigma_0^2}\right) \right]^+ & D > \sigma_0^2 \end{cases},
\end{aligned}$$

where $H_2(z) \equiv -z \log z - (1-z) \log(1-z)$ for $z \in [0, 1]$, $[x]^+ = \max\{0, x\}$ and

$$g(x_2) = \frac{\frac{\alpha}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x_2^2}{2\sigma_1^2}\right)}{\frac{\alpha}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x_2^2}{2\sigma_1^2}\right) + \frac{1-\alpha}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x_2^2}{2\sigma_0^2}\right)}. \quad (351)$$

Proof: The converse proof is straightforward by observing that when $D \leq \sigma_0^2$:

$$I(X_2; U|X_1) = h(X_2|X_1) - h(X_2|U, X_1) \geq h(X_2|X_1) - \frac{1}{2} \log(2\pi e D) \quad (352)$$

and $h(X_2|X_1) = \frac{\alpha}{2} \log(2\pi e \sigma_1^2) + \frac{1-\alpha}{2} \log(2\pi e \sigma_0^2)$. For the case when $\sigma_0^2 < D \leq \alpha \sigma_1^2 + (1-\alpha)\sigma_0^2$ we can write:

$$\begin{aligned}
I(X_1; U|X_1) &= h(X_2|X_1) - h(X_2|U, X_1) \\
&\geq h(X_2|X_1) - \alpha h(X_2|U, X_1 = 1) - (1-\alpha)h(X_2|X_1 = 0) \\
&= \frac{\alpha}{2} \log(2\pi e \sigma_1^2) - \alpha \log\left(\frac{2\pi e(D - (1-\alpha)\sigma_0^2)}{\alpha}\right) \quad (353)
\end{aligned}$$

$$= \frac{\alpha}{2} \log\left(\frac{\alpha \sigma_1^2}{D - (1-\alpha)\sigma_0^2}\right). \quad (354)$$

When $D > \alpha \sigma_1^2 + (1-\alpha)\sigma_0^2$ we can lower bound the mutual information by zero.

The achievability follows from the choice:

$$g(U, X_1) = \begin{cases} \left(\frac{\sigma_0^2}{\sigma_0^2 + \sigma_{Z_0}^2}\right) U & \text{if } X_1 = 0 \\ \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_{Z_1}^2}\right) U & \text{if } X_1 = 1, \end{cases} \quad (355)$$

and by setting the auxiliary random variable:

$$U = \begin{cases} X_2 + Z_0 & \text{if } X_1 = 0, \\ X_2 + Z_1 & \text{if } X_1 = 1, \end{cases} \quad (356)$$

where Z_0, Z_1 are zero-mean Gaussian random variables, independent from X_2 and X_1 and with variances given by:

$$\sigma_{Z_0}^2 = \frac{D\sigma_0^2}{\sigma_0^2 - D}, \quad \sigma_{Z_1}^2 = \frac{D\sigma_1^2}{\sigma_1^2 - D}, \quad (357)$$

for $D \leq \sigma_0^2$ while for $\sigma_0^2 < D \leq \alpha\sigma_1^2 + (1-\alpha)\sigma_0^2$, we choose:

$$\sigma_{Z_0}^2 \rightarrow \infty, \quad \sigma_{Z_1}^2 = \frac{[D - (1-\alpha)\sigma_0^2]\sigma_1^2}{\alpha\sigma_1^2 - [D - (1-\alpha)\sigma_0^2]}. \quad (358)$$

Finally, for $D > \alpha\sigma_1^2 + (1-\alpha)\sigma_0^2$, we let $\sigma_{Z_0}^2 \rightarrow \infty$ and $\sigma_{Z_1}^2 \rightarrow \infty$. ■

Unfortunately, the non-cooperative region is hard to evaluate for the assumed source model¹¹. In order to present some comparison between the cooperative and non-cooperative case let us fix the same value for the rate R_1 in both cases and compare the rate R_2 that can be obtained in each case. Clearly, in this way, we are not taking into account the gain in R_1 that could be obtained by the cooperative scheme (as $H(X_1|X_2) \leq H(X_1|U)$ for every $U \oplus X_2 \oplus X_1$). For both schemes, it follows that for fixed R_1 :

$$R_2 > \max \{I(X_2; U|X_1), H(X_1) + I(X_2; U|X_1) - R_1\}. \quad (359)$$

From Theorem 7, we can compute (359) for the cooperative case. For the non-cooperative case we need to obtain a lower bound on $I(X_2; U|X_1)$ for $p_{U|X_1} \in \mathcal{L}^*$. It is easy to check that:

$$I(X_2; U|X_1) \geq \frac{1-\alpha}{2} \log \left(\frac{\sigma_0^2}{\beta_0} \right) + \frac{\alpha}{2} \log \left(\frac{\sigma_1^2}{\beta_1} \right), \quad (360)$$

where

$$\beta_0 = \mathbb{E}_{X_2 U|X_1} \left[\left(X_2 - \mathbb{E}_{X_2|U, X_1=0} [X_2|U, X_1=0] \right)^2 | X_1=0 \right], \quad (361)$$

$$\beta_1 = \mathbb{E}_{X_2 U|X_1} \left[\left(X_2 - \mathbb{E}_{X_2|U, X_1=1} [X_2|U, X_1=1] \right)^2 | X_1=1 \right]. \quad (362)$$

The distortion constraint imposes the condition:

$$(1-\alpha)\beta_0 + \alpha\beta_1 \leq D. \quad (363)$$

¹¹However, there are cases where an exact characterization is possible. This is the case, for example, when X_1 and X_2 are the input and output of binary channel with crossover probability α and the distortion function is the Hamming distance [5].

In order to guarantee that (360) and (361) are achievable, under the Markov constraint $U \rightarrow X_2 \rightarrow X_1$, the following conditions on $p_{U|X_2}(u|x_2)$ should be satisfied:

$$\frac{p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x_2^2}{2\sigma_0^2}\right)}{\int_{-\infty}^{\infty} p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x_2^2}{2\sigma_0^2}\right) dx_2} = \frac{1}{\sqrt{2\pi}\beta_0} \exp\left\{-\frac{(x_2 - f_0(u))^2}{2\beta_0}\right\} \quad (364)$$

and

$$\frac{p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x_2^2}{2\sigma_1^2}\right)}{\int_{-\infty}^{\infty} p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x_2^2}{2\sigma_1^2}\right) dx_2} = \frac{1}{\sqrt{2\pi}\beta_1} \exp\left\{-\frac{(x_2 - f_1(u))^2}{2\beta_1}\right\}, \quad (365)$$

where

$$f_0(U) \equiv \mathbb{E}_{X_2|U, X_1=0} [X_2|U, X_1=0], \quad f_1(U) \equiv \mathbb{E}_{X_2|U, X_1=1} [X_2|U, X_1=1]. \quad (366)$$

The characterization of all distributions $p_{U|X_2}(u|x_2)$ that satisfies (365) and (366) appears to be a difficult problem. In order to show a numerical example, we shall simply assume that:

$$p_{U|X_2}(u|x_2) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left\{-\frac{(u - x_2)^2}{2\sigma_w^2}\right\}. \quad (367)$$

Indeed, this choice satisfies simultaneously expressions (365) and (366). In this way, we can calculate the corresponding values of β_0 and β_1 obtaining the parametrization of $I(X_2; U|X_1)$ as function of σ_w^2 :

$$I(X_2; U|X_1) = \frac{1-\alpha}{2} \log\left(\frac{\sigma_0^2 + \sigma_w^2}{\sigma_w^2}\right) + \frac{\alpha}{2} \log\left(\frac{\sigma_1^2 + \sigma_w^2}{\sigma_w^2}\right) \quad (368)$$

with the following constraint:

$$(1-\alpha) \frac{\sigma_w^2 \sigma_0^2}{\sigma_0^2 + \sigma_w^2} + \alpha \frac{\sigma_w^2 \sigma_1^2}{\sigma_1^2 + \sigma_w^2} = D. \quad (369)$$

We can replace (369) in (359) to obtain an indication of the performance of the non-cooperative case when R_1 is fixed.

We present now some numerical evaluations. As equation (359) is valid for both the cooperative and the non-cooperative setups, it is sufficient to compare the mutual information term $I(X_2; U|X_1)$ for each of them. Let us consider the next scenarios:

- 1) $\alpha = 0.1, \sigma_0^2 = 0.01, \sigma_1^2 = 2,$
- 2) $\alpha = 0.1, \sigma_0^2 = 0.5, \sigma_1^2 = 2.$

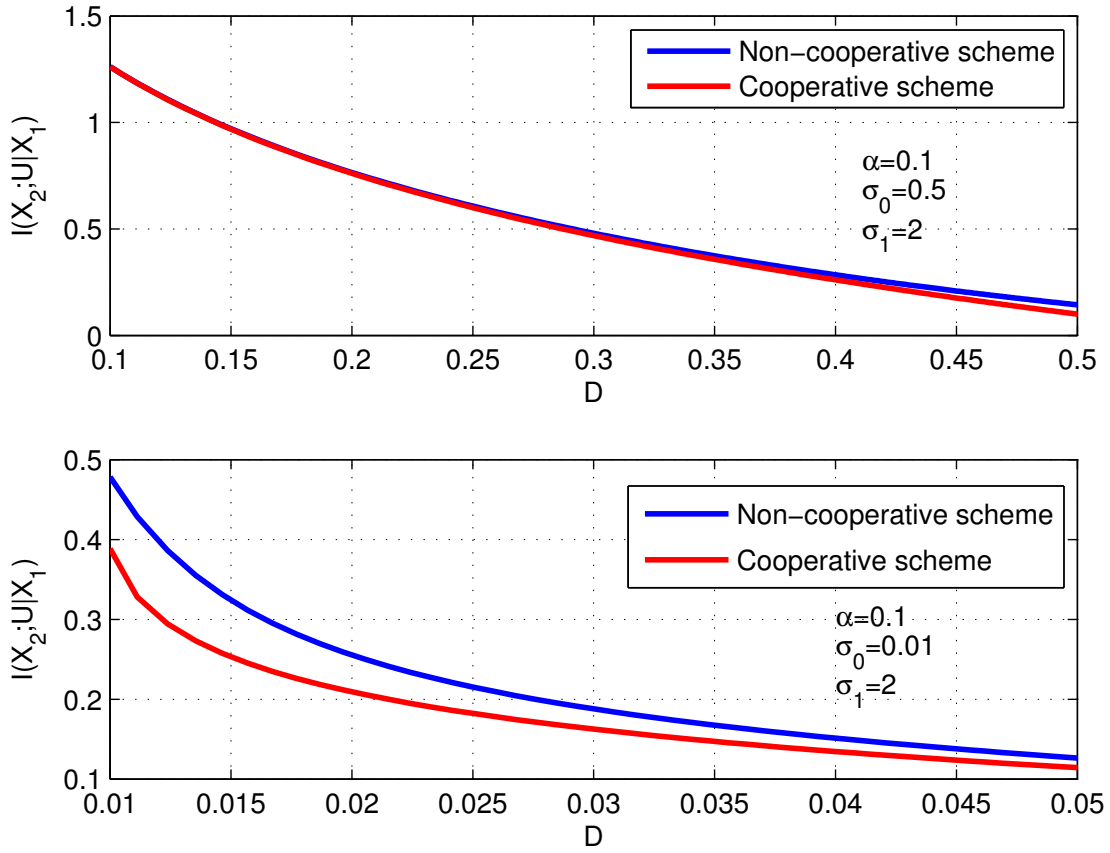


Figure 7: Comparison between the cooperative and the non-cooperative schemes.

From Fig. 7 we see that in the case $\sigma_0^2 \ll \sigma_1^2$ the gain of the cooperative scheme is pretty noticeable. However, as σ_0^2 becomes comparable to σ_1^2 the gains are reduced. This was expected from the fact that as $\sigma_0^2 \rightarrow \sigma_1^2$, the random variable X_2 converges to a Gaussian distribution. In that case, the reconstruction of X_2 at Node 3 is equivalent for the cooperative scenario to a lossy source coding problem with side information X_1 at both the encoder and the decoder while for the non-cooperative setting to the standard Wyner-Ziv problem. It is known that in this case there is no gains that can be expected [2].

B. Interactive Lossless Source Coding

Consider now the problem described in Fig. 8 where encoder 1 wishes to communicate *lossless* the source X_1^n to two decoders which observe sources X_2^n and X_3^n . At the same time node 1

wishes to recover X_2^n and X_3^n lossless. Similarly the other encoders want to communicate *lossless* their sources and recover the sources from the rest. It is wanted to do this through K rounds of exchanges.

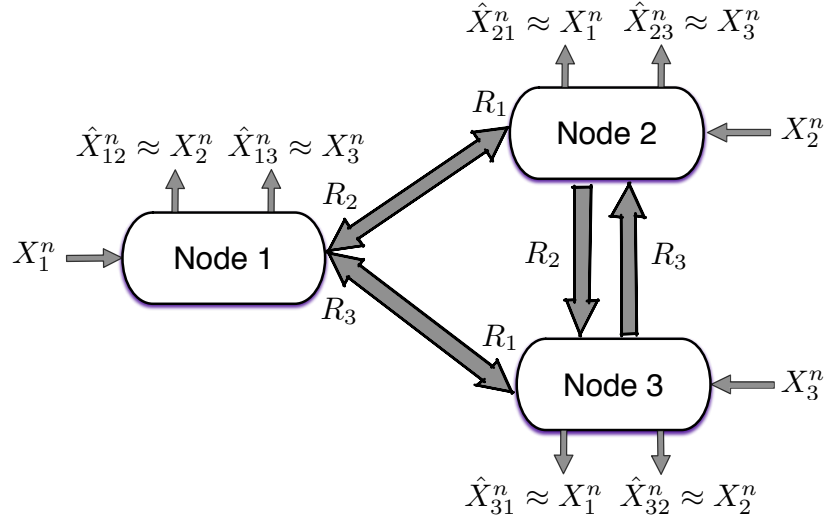


Figure 8: Interactive lossless source coding.

Theorem 8 (Interactive lossless source coding): The rate region of the setting described in Fig. 8 is given by the set of all tuples satisfying:

$$R_1 > H(X_1 | X_2 X_3) , \quad (370)$$

$$R_2 > H(X_2 | X_1 X_3) , \quad (371)$$

$$R_3 > H(X_3 | X_1 X_2) , \quad (372)$$

$$R_1 + R_2 > H(X_1 X_2 | X_3) , \quad (373)$$

$$R_1 + R_3 > H(X_1 X_3 | X_2) , \quad (374)$$

$$R_2 + R_3 > H(X_2 X_3 | X_1) . \quad (375)$$

Remark 16: It is worth observing that the multiple exchanges of descriptions between all nodes cannot increase the rate compared to standard Slepian-Wolf coding [1].

Proof: The achievability part is a standard exercise. The converse proof is straightforward from cut-set arguments. For these reasons both proofs are omitted. ■

We should note that for this important case, Theorem 1 does not provide the optimal rate-region. That is, the coding scheme used is not optimal for this case. In fact, from Theorem 1 and for this problem we can obtain the following achievable region¹²:

$$R_1 > H(X_1|X_2) , \quad (376)$$

$$R_2 > H(X_2|X_1X_3) , \quad (377)$$

$$R_3 > H(X_3|X_1X_2) , \quad (378)$$

$$R_1 + R_2 > H(X_1X_2|X_3) , \quad (379)$$

$$R_2 + R_3 > H(X_2X_3|X_1) . \quad (380)$$

It is easily seen that in this region the node 2 is not performing joint decoding of the descriptions generated at node 1 and 3. Because of the encoding ordering assumed ($1 \rightarrow 2 \rightarrow 3$) and the fact that the common description generated in node 2 should be conditionally generated on the common description generated at node 1, node 2 has to recover first this common description first. At the end, it recovers the common description generated at node 3. On the other hand, nodes 1 and 3 perform joint decoding of the common information generated at nodes 2 and 3, and at nodes 1 and 2, respectively. Clearly, this is consequence of the *sequential* encoding and decoding structure imposed between the nodes in the network and which is the basis of the interaction.

If all nodes would be allowed to perform a joint decoding procedure in order to recover all the exchanged descriptions only at the end of each round, this problem would not appear. However, this would destroy the sequential encoding-decoding structure assumed in the paper which seems to be optimal in other situations.

VII. SUMMARY

The three-node multiterminal lossy source coding problem was investigated. This problem is not a straightforward generalization of the original problem posed by Kaspi in 1985. As this general problem encompasses several open problems in multiterminal rate distortion the

¹²Consider $U_{1 \rightarrow 23,1} \equiv X_1$, $U_{2 \rightarrow 13,1} \equiv X_2$ and $U_{3 \rightarrow 12,1} \equiv X_3$ and the other auxiliary random variables to be constants for all $l \in [1 : K]$.

mathematical complexity of it is formidable. For that reason we only provided a general inner bound for the rate distortion region. It is shown that this (rather involved) inner bound contains several rate-distortion regions of some relevant source coding settings. In this way, besides the non-trivial extension of the interactive two terminal problem, our results can be seen as a generalization and hence unification of several previous works in the field. We also showed, that our inner bound provides definite answers to special cases of the general problem. It was shown that in some cases the cooperation induced by the interaction can be helpful while in others not. It is clear that further study is needed on the topic of multiple terminal cooperative coding, including a proper generalization to larger networks and to the problem of interactively estimating arbitrary functions of the sources sensed at the nodes.

APPENDIX A

STRONGLY TYPICAL SEQUENCES AND RELATED RESULTS

In this appendix we introduce standard notions in information theory but suited for the mathematical developments and proof needed in this work. The results presents can be easily derived from the standard formulations provided in [22] and [23]. Be \mathcal{X} and \mathcal{Y} finite alphabets and $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. With $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we denote the set of all probability distributions on $\mathcal{X} \times \mathcal{Y}$. We define the *strongly δ -typical* sets as:

Definition 3 (Strongly typical set): Consider $p \in \mathcal{P}(\mathcal{X})$ and $\delta > 0$. We say that $x^n \in \mathcal{X}^n$ is $p\delta$ - strongly typical if $x^n \in \mathcal{T}_{[p]\delta}^n$ with:

$$\mathcal{T}_{[p]\delta}^n = \left\{ x^n \in \mathcal{X}^n : \left| \frac{N(a|x^n)}{n} - p(a) \right| \leq \frac{\delta}{\|\mathcal{X}\|}, \forall a \in \mathcal{X} \text{ such that } p(a) \neq 0 \right\}, \quad (381)$$

where $N(a|x^n)$ denotes de number of occurrences of $a \in \mathcal{X}$ in x^n and $p \in \mathcal{P}(\mathcal{X})$. When $X \sim p_X(x)$ we can denote the corresponding set of strongly typical sequences as $\mathcal{T}_{[X]\delta}^n$.

Similarly, given $p_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we can construct the set of δ -jointly typical sequences as:

$$\begin{aligned} \mathcal{T}_{[XY]\delta}^n = \left\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| \frac{N(a, b|x^n, y^n)}{n} - p_{XY}(a, b) \right| \leq \frac{\delta}{\|\mathcal{X}\|\|\mathcal{Y}\|}, \right. \\ \left. \forall (a, b) \in \mathcal{X} \times \mathcal{Y} \text{ such that } p_{XY}(a, b) \neq 0 \right\}. \end{aligned} \quad (382)$$

We also define the *conditional* typical sequences. In precise terms, given $x^n \in \mathcal{X}^n$ we consider the set:

$$\begin{aligned} \mathcal{T}_{[Y|X]\delta}^n(x^n) = \left\{ y^n \in \mathcal{Y}^n : \left| \frac{N(a, b|x^n, y^n)}{n} - p_{XY}(a, b) \right| \leq \frac{\delta}{\|\mathcal{X}\|\|\mathcal{Y}\|}, \right. \\ \left. \forall (a, b) \in \mathcal{X} \times \mathcal{Y} \text{ such that } p_{XY}(a, b) \neq 0 \right\}. \end{aligned} \quad (383)$$

Notice that we the following is an alternative writing of this set:

$$\mathcal{T}_{[Y|X]\delta}^n(x^n) = \{ y^n \in \mathcal{Y}^n : (x^n, y^n) \in \mathcal{T}_{[XY]\delta}^n \}. \quad (384)$$

We have several useful and standard Lemmas, which will be presented without proof (except for the last one):

Lemma 1 (Properties of typical sets [23]): The following are true:

- 1) Consider $(x^n, y^n) \in \mathcal{T}_{[XY]\epsilon}^n$. Then, $x^n \in \mathcal{T}_{[X]\epsilon}^n$, $y^n \in \mathcal{T}_{[Y]\epsilon}^n$, $x^n \in \mathcal{T}_{[X|Y]\epsilon}^n(y^n)$ and $y^n \in \mathcal{T}_{[Y|X]\epsilon}^n(x^n)$.

- 2) Be $T_{[Y|X]\epsilon}^n(x^n)$ with $x^n \notin \mathcal{T}_{[X]\epsilon}^n$. Then $T_{[Y|X]\epsilon}^n(x^n) = \emptyset$.
- 3) Be $(X^n, Y^n) \sim \prod_{t=1}^n p_{XY}(x_t, y_t)$. If $x^n \in \mathcal{T}_{[X]\epsilon}^n$ we have

$$2^{-n(H(X)+\delta(\epsilon))} \leq p_{X^n}(x^n) \leq 2^{-n(H(X)-\delta(\epsilon))}$$

with $\delta(\epsilon) \rightarrow 0$ when $\epsilon \rightarrow 0$. Similarly, if $y^n \in \mathcal{T}_{[Y|X]\epsilon}^n(x^n)$:

$$2^{-n(H(Y|X)+\delta'(\epsilon))} \leq p_{Y^n|X^n}(y^n|x^n) \leq 2^{-n(H(Y|X)-\delta'(\epsilon))}$$

with $\delta'(\epsilon) \rightarrow 0$ when $\epsilon \rightarrow 0$.

Lemma 2 (Conditional typicality lemma [23]): Consider the product measure $\prod_{t=1}^n p_{XY}(x_t, y_t)$.

Using that measure, we have the following

$$\Pr \{ \mathcal{T}_{[X]\epsilon}^n \} \geq 1 - \mathcal{O} \left(c_1^{-nf(\epsilon)} \right), \quad c_1 > 1$$

where $f(\epsilon) \rightarrow 0$ when $\epsilon \rightarrow 0$. In addition, for every $x^n \in \mathcal{T}_{[X]\epsilon'}^n$ with $\epsilon' < \frac{\epsilon}{\|\mathcal{Y}\|}$ we have:

$$\Pr \{ \mathcal{T}_{[Y|X]\epsilon}^n(x^n) \} \geq 1 - \mathcal{O} \left(c_2^{-ng(\epsilon, \epsilon')} \right), \quad c_2 > 1$$

where $g(\epsilon, \epsilon') \rightarrow 0$ when $\epsilon, \epsilon' \rightarrow 0$.

Lemma 3 (Size of typical sets [23]): Given $p_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we have

$$\frac{1}{n} \|\mathcal{T}_{[X]\epsilon}^n\| \leq H(X) + \delta(\epsilon), \quad \frac{1}{n} \|\mathcal{T}_{[X]\epsilon}^n\| \geq H(X) - \delta'(\epsilon, n)$$

where $\delta(\epsilon), \delta'(\epsilon, n) \rightarrow 0$ when $\epsilon \rightarrow 0$ and $n \rightarrow \infty$. Similarly for every $x^n \in \mathcal{X}^n$ we have:

$$\frac{1}{n} \|\mathcal{T}_{[Y|X]\epsilon}^n(x^n)\| \leq H(Y|X) + \delta(\epsilon)$$

with $\delta(\epsilon) \rightarrow 0$ when $\epsilon \rightarrow 0$. In addition, for every $x^n \in \mathcal{T}_{[X]\epsilon'}^n$ with $\epsilon' < \frac{\epsilon}{\|\mathcal{Y}\|}$ we have:

$$\frac{1}{n} \|\mathcal{T}_{[Y|X]\epsilon}^n(x^n)\| \geq H(Y|X) - \delta'(\epsilon, \epsilon', n)$$

where $\delta'(\epsilon, \epsilon', n) \rightarrow 0$ when $\epsilon, \epsilon' \rightarrow 0$ and $n \rightarrow \infty$.

Lemma 4 (Joint typicality lemma [23]): Consider $(X, Y, Z) \sim p_{XYZ}(x, y, z)$ and $(x^n, y^n) \in \mathcal{T}_{[XY]\epsilon'}^n$ with $\epsilon' < \frac{\epsilon}{\|\mathcal{Z}\|}$ and $\epsilon < \epsilon''$. If $p_{Z^n|X^n}(z^n|x^n) = \frac{\mathbb{1}_{\{z^n \in \mathcal{T}_{[Z|X]\epsilon''}^n(x^n)\}}}{\|\mathcal{T}_{[Z|X]\epsilon''}^n(x^n)\|}$ there exists $\delta'(\epsilon, \epsilon', \epsilon'', n)$ which goes to zero as $\epsilon, \epsilon', \epsilon'' \rightarrow 0$ and $n \rightarrow \infty$ and:

$$2^{-n(I(Y;Z|X)+\delta')} \leq \Pr \left\{ \tilde{Z}^n \in \mathcal{T}_{[Z|XY]\epsilon}^n(x^n, y^n) \right\} \leq 2^{-n(I(Y;Z|X)-\delta')}$$

Lemma 5 (Covering Lemma [22]): Be $(U, V, X) \sim p_{UVX}$ and $(x^n, u^n) \in \mathcal{T}_{[XU]\epsilon'}^n$, $\epsilon' < \frac{\epsilon}{\|\mathcal{V}\|}$ and $\epsilon < \epsilon''$. Consider also $\{V^n(m)\}_{m=1}^{2^{nR}}$ random vectors which are independently generated according to $\frac{\mathbb{1}\{v^n \in \mathcal{T}_{[V|U]\epsilon''}^n(u^n)\}}{\|\mathcal{T}_{[V|U]\epsilon''}^n(u^n)\|}$. Then:

$$\Pr \{V^n(m) \notin \mathcal{T}_{[V|U]\epsilon}^n(x^n, u^n) \text{ for all } m\} \xrightarrow{n \rightarrow \infty} 0 \quad (385)$$

uniformly for every $(x^n, u^n) \in \mathcal{T}_{[XU]\epsilon'}^n$ if:

$$R > I(V; X|U) + \delta(\epsilon, \epsilon', \epsilon'', n) \quad (386)$$

where $\delta(\epsilon, \epsilon', \epsilon'', n) \rightarrow 0$ when $\epsilon, \epsilon', \epsilon'' \rightarrow 0$ and $n \rightarrow \infty$.

Corollary 1: Assume the conditions in Lemma 5, and also:

$$\Pr \{(X^n, U^n) \in \mathcal{T}_{[XU]\epsilon'}^n\} \xrightarrow{n \rightarrow \infty} 1. \quad (387)$$

Then:

$$\Pr \{(U^n, X^n, V^n(m)) \notin \mathcal{T}_{[UXV]\epsilon}^n \text{ for all } m\} \xrightarrow{n \rightarrow \infty} 0 \quad (388)$$

when (386) is satisfied.

Lemma 6 (Packing Lemma [22]): Be $(U_1 U_2 W V_1 V_2 X) \sim p_{U_1 U_2 W V_1 V_2 X}$, $(x^n, w^n, v_1^n, v_2^n) \in \mathcal{T}_{[XWV_1V_2]\epsilon'}^n$ and $\epsilon' < \frac{\epsilon}{\|\mathcal{U}_1\| \|\mathcal{U}_2\|}$ and $\epsilon < \min\{\epsilon_1, \epsilon_2\}$. Consider random vectors $\{U_1^n(m_1)\}_{m_1=1}^{\mathcal{A}_1}$ and $\{U_2^n(m_2)\}_{m_2=1}^{\mathcal{A}_2}$ which are independently generated according to

$$\frac{\mathbb{1}\{u_i^n \in \mathcal{T}_{[U_i|V_iW]\epsilon_i}^n(v_i^n, w^n)\}}{\|\mathcal{T}_{[U_i|V_iW]\epsilon_i}^n(w^n, v_i^n)\|}, \quad i = 1, 2,$$

and $\mathcal{A}_1, \mathcal{A}_2$ are positive random variables independent of everything else. Then

$$\Pr \{(U_1^n(m_1), U_2^n(m_2)) \in \mathcal{T}_{[U_1U_2|XWV_1V_2]\epsilon}^n(x^n, w^n, v_1^n, v_2^n) \text{ for some } (m_1, m_2)\} \xrightarrow{n \rightarrow \infty} 0 \quad (389)$$

uniformly for every $(x^n, w^n, v_1^n, v_2^n) \in \mathcal{T}_{[XWV_1V_2]\epsilon'}^n$ provided that:

$$\frac{\log \mathbb{E}[\mathcal{A}_1 \mathcal{A}_2]}{n} < I(U_1; X V_2 U_2 | W V_1) + I(U_2; X V_1 U_1 | W V_2) - I(U_1; U_2 | X W V_1 V_2) - \delta \quad (390)$$

where $\delta \equiv \delta(\epsilon, \epsilon', \epsilon_1, \epsilon_2, n) \rightarrow 0$ when $\epsilon, \epsilon', \epsilon_1, \epsilon_2 \rightarrow 0$ and $n \rightarrow \infty$.

Corollary 2: Assume the conditions in Lemma 6, and also:

$$\Pr \{(X^n, W^n, V_1^n, V_2^n) \in \mathcal{T}_{[XWV_1V_2]\epsilon'}^n\} \xrightarrow{n \rightarrow \infty} 1 \quad (391)$$

Then:

$$\Pr \left\{ (U_1^n(m_1), U_2^n(m_2), X^n, W^n, V_1^n, V_2^n) \in \mathcal{T}_{[U_1 U_1 X W V_1 V_2]^\epsilon}^n \text{ for some } (m_1, m_2) \right\} \xrightarrow{n \rightarrow \infty} 0 \quad (392)$$

when (390) is satisfied.

Lemma 7 (Generalized Markov Lemma [24]): Consider a pmf $p_{U_{XY}}$ belonging to $\mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ and that satisfies de following:

$$Y \ominus X \ominus U$$

Consider $(x^n, y^n) \in \mathcal{T}_{[XY]^\epsilon}^n$ and random vectors U^n generated according to:

$$\Pr \left\{ U^n = u^n \middle| x^n, y^n, U^n \in \mathcal{T}_{[U|X]^\epsilon}^n(x^n) \right\} = \frac{\mathbb{1} \left\{ u^n \in \mathcal{T}_{[U|X]^\epsilon}^n(x^n) \right\}}{\|\mathcal{T}_{[U|X]^\epsilon}^n(x^n)\|} \quad (393)$$

For sufficiently small $\epsilon, \epsilon', \epsilon''$ the following holds uniformly for every $(x^n, y^n) \in \mathcal{T}_{[XY]^\epsilon}^n$:

$$\Pr \left\{ U^n \notin \mathcal{T}_{[U|XY]^\epsilon}^n(x^n, y^n) \middle| x^n, y^n, U^n \in \mathcal{T}_{[U|X]^\epsilon}^n(x^n) \right\} = \mathcal{O}(c^{-n}) \quad (394)$$

where $c > 1$.

Corollary 3: Assume the conditions in Lemma 7, and also:

$$\Pr \left\{ (X^n, Y^n) \in \mathcal{T}_{[XY]^\epsilon}^n \right\} \xrightarrow{n \rightarrow \infty} 1 \quad (395)$$

and that uniformly for every $(x^n, y^n) \in \mathcal{T}_{[XY]^\epsilon}^n$:

$$\Pr \left\{ U^n \notin \mathcal{T}_{[U|X]^\epsilon}^n(x^n) \middle| x^n, y^n \right\} \xrightarrow{n \rightarrow \infty} 0 \quad (396)$$

we obtain:

$$\Pr \left\{ (U^n, X^n, Y^n) \in \mathcal{T}_{[UXY]^\epsilon}^n \right\} \xrightarrow{n \rightarrow \infty} 1. \quad (397)$$

Lemma 7 and Corollary 3 will be central for us. They will guarantee the joint typicality of the descriptions generated in different encoders considering the pmf of the chosen descriptions induced by the coding scheme used. The original proof of this result is given in [4] and involves a combination of rather sophisticated algebraic and combinatorial arguments over finite alphabets. Alternative proof was also provided in [22],

$$\Pr \left\{ (U^n, X^n, Y^n) \in \mathcal{T}_{[UXY]^\epsilon}^n \right\} \xrightarrow{n \rightarrow \infty} 1 \quad (398)$$

which strongly relies on a rather obscure result by Uhlmann [25] on combinatorics. In [24] a short and more general proof of this result is given.

We next present a result which will be useful for proving Theorem 1. In order to use the Markov Lemma we need to show that the descriptions induced by the encoding procedure in each node satisfies (393).

Lemma 8 (Encoding induced distribution): Consider a pmf p_{UXW} belonging to $\mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{W})$ and $\epsilon' \geq \epsilon$. Be $\{U^n(m)\}_{m=1}^S$ random vectors independently generated according to

$$\frac{\mathbb{1} \left\{ u^n \in \mathcal{T}_{[U|W]\epsilon'}^n(w^n) \right\}}{\|\mathcal{T}_{[U|W]\epsilon'}^n(w^n)\|}$$

and where (W^n, X^n) are generated with an arbitrary distribution. Once these vectors are generated, and given x^n and w^n , we choose one of them if:

$$(u^n(m), w^n, x^n) \in \mathcal{T}_{[UXW]\epsilon}^n, \text{ for some } m \in [1 : S] . \quad (399)$$

If there are various vectors u^n that satisfies this we choose the one with smallest index. If there are none we choose an arbitrary one. Let M denote the index chosen. Then we have that:

$$\Pr \left\{ U^n(M) = u^n \middle| x^n, w^n, U^n(M) \in \mathcal{T}_{[UXW]\epsilon}^n(x^n, w^n) \right\} = \frac{\mathbb{1} \left\{ u^n \in \mathcal{T}_{[UXW]\epsilon}^n(x^n, w^n) \right\}}{\|\mathcal{T}_{[UXW]\epsilon}^n(x^n, w^n)\|} . \quad (400)$$

Proof: From the selection procedure for M we know that:

$$M = f \left(\mathbb{1} \left\{ (U^n(m), X^n, W^n) \in \mathcal{T}_{[UXW]\epsilon}^n \right\}, m \in [1 : S] \right) , \quad (401)$$

where $f(\cdot)$ is an appropriate function. Moreover, because of this and the way in which the random vectors U^n are generated we have:

$$U^n(M) \ominus (\mathbb{1} \left\{ (U^n(M), X^n, W^n) \in \mathcal{T}_{[UXW]\epsilon}^n \right\}, W^n, X^n) \ominus M . \quad (402)$$

We can write:

$$\begin{aligned} & \Pr \left\{ U^n(M) = u^n \middle| x^n, w^n, U^n(M) \in \mathcal{T}_{[UXW]\epsilon}^n(x^n, w^n) \right\} = \\ & \sum_{m=1}^S \Pr \left\{ M = m \middle| x^n, w^n, U^n(m) \in \mathcal{T}_{[UXW]\epsilon}^n(x^n, w^n) \right\} \times \\ & \Pr \left\{ U^n(m) = u^n \middle| x^n, w^n, U^n(m) \in \mathcal{T}_{[UXW]\epsilon}^n(x^n, w^n), M = m \right\} . \end{aligned} \quad (403)$$

From (402), the second probability term in the RHS of (403) can be written as:

$$\Pr \left\{ U^n(m) = u^n \middle| x^n, w^n, U^n(m) \in \mathcal{T}_{[U|XW]\epsilon}^n(x^n, w^n) \right\} \quad \forall m. \quad (404)$$

We are going to analyze this term. It is clear that we can write:

$$\begin{aligned} \Pr \left\{ U^n(m) = u^n, U^n(m) \in \mathcal{T}_{[U|XW]\epsilon}^n(x^n, w^n) \middle| x^n, w^n \right\} &= \mathbb{1} \left\{ u^n \in \mathcal{T}_{[U|WX]\epsilon}^n(x^n, w^n) \right\} \\ &\quad \times \Pr \left\{ U^n(m) = u^n \middle| x^n, w^n \right\} \quad \forall m \end{aligned} \quad (405)$$

■

This means that

$$\begin{aligned} &\Pr \left\{ U^n(m) = u^n \middle| x^n, w^n, U^n(m) \in \mathcal{T}_{[U|XW]\epsilon}^n(x^n, w^n) \right\} \\ &= \frac{\mathbb{1} \left\{ u^n \in \mathcal{T}_{[U|WX]\epsilon}^n(x^n, w^n) \cap \mathcal{T}_{[U|W]\epsilon'}^n(w^n) \right\}}{\Pr \left\{ U^n(m) \in \mathcal{T}_{[U|XW]\epsilon}^n(x^n, w^n) \middle| x^n, w^n \right\} \|\mathcal{T}_{[U|W]\epsilon'}^n(w^n)\|} \quad \forall m. \end{aligned} \quad (406)$$

From (406) and the fact that for $\epsilon' \geq \epsilon$, we have that $\mathcal{T}_{[U|XW]\epsilon}^n(x^n, w^n) \subseteq \mathcal{T}_{[U|W]\epsilon'}^n(w^n)$ we obtain:

$$\Pr \left\{ U^n(m) = u^n \middle| x^n, w^n, U^n(m) \in \mathcal{T}_{[U|XW]\epsilon}^n(x^n, w^n) \right\} = \frac{\mathbb{1} \left\{ u^n \in \mathcal{T}_{[U|WX]\epsilon}^n(x^n, w^n) \right\}}{\|\mathcal{T}_{[U|XW]\epsilon}^n(x^n, w^n)\|} \quad \forall m. \quad (407)$$

From this equation and (403) we easily obtain the desired result.

We present, without proof, a useful result about reconstruction functions for lossy source coding problems:

Lemma 9 (Reconstruction functions for degraded random variables [13]): Consider random variables (X, Y, Z) such that $X \dashv Y \dashv Z$. Consider an arbitrary function $\hat{X} = f(Y, Z)$ and an arbitrary positive distortion function $d(\cdot, \cdot)$. Then $\exists g^*(Y)$ such that

$$E[d(X, g^*(Y))] \leq E[d(X, f(Y, Z))] . \quad (408)$$

Finally we present two lemmas about Markov chains induced by the interactive encoding schemes which will be relevant for the paper converse results

Lemma 10 (Markov chains induced by interactive encoding of two nodes): Consider a set of three sources $(X^n, Y^n, Z^n) \sim \prod_{t=1}^n p_{XYZ}(x_t, y_t, z_t)$ and integer $K \in \mathbb{N}$. For each $l \in [1 : K]$ consider arbitrary message sets $\mathcal{I}_x^l, \mathcal{I}_y^l$ and arbitrary functions

$$f_x^l(X^n, \mathcal{I}_x^{[1:l-1]}, \mathcal{I}_y^{[1:l-1]}) = \mathcal{I}_x^l, \quad (409)$$

$$f_y^l(Y^n, \mathcal{I}_x^{[1:l]}, \mathcal{I}_y^{[1:l-1]}) = \mathcal{I}_y^l \quad (410)$$

with $\mathcal{J}_x^l \in \mathcal{I}_x^l$ and $\mathcal{J}_y^l \in \mathcal{I}_y^l$. The following Markov chain relations are valid for each $t \in [1 : n]$ and $l \in [1 : K]$:

- 1) $(\mathcal{J}_x^1, X_{[1:t-1]}, Y_{[t+1:n]}) \ominus X_{[t]} \ominus (Y_{[t]}, Z_{[t]})$,
- 2) $(\mathcal{J}_x^l, X_{[t+1:n]}) \ominus (\mathcal{J}_x^{[1:l-1]}, \mathcal{J}_y^{[1:l-1]}, X_{[1:t]}, Y_{[t+1:n]}) \ominus (Y_{[t]}, Z_{[t]})$,
- 3) $(\mathcal{J}_y^l, Y_{[1:t-1]}) \ominus (\mathcal{J}_x^{[1:l]}, \mathcal{J}_y^{[1:l-1]}, X_{[1:t-1]}, Y_{[t:n]}) \ominus (X_{[t]}, Z_{[t]})$,
- 4) $X_{[t+1:n]} \ominus (\mathcal{J}_x^{[1:K]}, \mathcal{J}_y^{[1:K]}, X_{[1:t]}, Y_{[t+1:n]}) \ominus (Y_{[t]}, Z_{[t]})$,
- 5) $Y_{[1:t-1]} \ominus (\mathcal{J}_x^{[1:K]}, \mathcal{J}_y^{[1:K]}, X_{[1:t-1]}, Y_{[t:n]}) \ominus (X_{[t]}, Z_{[t]})$,
- 6) $(\mathcal{J}_x^{[1:K]}, \mathcal{J}_y^{[1:K]}, X_{[1:t-1]}, X_{[t+1:n]}, Z_{[1:t-1]}, Z_{[t+1:n]}, Y_{[1:t-1]}) \ominus (X_{[t]}, Y_{[t]}) \ominus Z_{[t]}$.

Proof: Relations 1), 2) and 3) were obtained in [13]. For completeness we present here a short proof of 2). The proof of 1) and 3) are similar. For simplicity let us consider $A = I(\mathcal{J}_x^l X_{[t+1:n]}; Y_{[t]} Z_{[t]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-1]} X_{[1:t]} Y_{[t+1:n]})$. We can write the following:

$$\begin{aligned}
A &\stackrel{(a)}{\leq} I(\mathcal{J}_x^l X_{[t+1:n]}; Y_{[1:t]} Z_{[t]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-1]} X_{[1:t]} Y_{[t+1:n]}) \\
&\stackrel{(b)}{=} I(X_{[t+1:n]}; Y_{[1:t]} Z_{[t]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-1]} X_{[1:t]} Y_{[t+1:n]}) \\
&= H(X_{[t+1:n]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-1]} X_{[1:t]} Y_{[t+1:n]}) - H(X_{[t+1:n]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-1]} X_{[1:t]} Z_{[t]} Y^n) \\
&\stackrel{(c)}{\leq} I(X_{[t+1:n]}; Y_{[1:t]} Z_{[t]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-2]} X_{[1:t]} Y_{[t+1:n]}) \\
&= H(Y_{[1:t]} Z_{[t]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-2]} X_{[1:t]} Y_{[t+1:n]}) - H(Y_{[1:t]} Z_{[t]} | \mathcal{J}_x^{[1:l-1]} \mathcal{J}_y^{[1:l-2]} X^n Z_{[t]} Y_{[t+1:n]}) \\
&\stackrel{(d)}{\leq} I(X_{[t+1:n]}; Y_{[1:t]} Z_{[t]} | \mathcal{J}_x^{[1:l-2]} \mathcal{J}_y^{[1:l-2]} X_{[1:t]} Y_{[t+1:n]}) \tag{411}
\end{aligned}$$

where

- step (a) follows from non-negativity of mutual information,
- step (b) follows from the fact that $\mathcal{J}_x^l = f_x^l(X^n, \mathcal{J}_x^{[1:l-1]}, \mathcal{J}_y^{[1:l-1]})$,
- step (c) follows from the fact that $\mathcal{J}_y^{l-1} = f_y^l(Y^n, \mathcal{J}_x^{[1:l-1]}, \mathcal{J}_y^{[1:l-2]})$ and that conditioning reduces entropy,
- step (d) follows from the fact that $\mathcal{J}_x^{l-1} = f_x^l(X^n, \mathcal{J}_x^{[1:l-2]}, \mathcal{J}_y^{[1:l-2]})$ and that conditioning reduces entropy.

Continuing this procedure we obtain:

$$A \leq I(X_{[t+1:n]}; Y_{[1:t]} Z_{[t]} | X_{[1:t]} Y_{[t+1:n]}) = 0. \tag{412}$$

This shows that 2) is true. 4) and 5) are straightforward consequences of 2) and 3). Just consider $\mathcal{J}_x^{K+1} = \mathcal{J}_y^{K+1} = \emptyset$. The proof for 6) is straightforward from the fact that $(\mathcal{J}_x^{[1:K]}, \mathcal{J}_y^{[1:K]})$ is only function of (X^n, Y^n) . ■

Corollary 4: Consider the setting in Lemma 10 with the following modifications:

- $X \ominus Z \ominus Y$,
- $f_x^l(X^n, Z^n, \mathcal{J}_x^{[1:l-1]}, \mathcal{J}_y^{[1:l-1]}) = \mathcal{J}_x^l$.

The following are true:

- 1) $(\mathcal{J}_x^1, Z_{[1:t-1]}, Y_{[t+1:n]}) \ominus Z_{[t]} \ominus Y_{[t]}$,
- 2) $(\mathcal{J}_x^l, Z_{[t+1:n]}) \ominus (\mathcal{J}_x^{[1:l-1]}, \mathcal{J}_y^{[1:l-1]}, Z_{[1:t]}, Y_{[t+1:n]}) \ominus Y_{[t]}$,
- 3) $(\mathcal{J}_y^l, Y_{[1:t-1]}) \ominus (\mathcal{J}_x^{[1:l]}, \mathcal{J}_y^{[1:l-1]}, Z_{[1:t-1]}, Y_{[t:n]}) \ominus (X_{[t]}, Z_{[t]})$,
- 4) $(Z_{[t+1:n]}, X^n) \ominus (\mathcal{J}_x^{[1:K]}, \mathcal{J}_y^{[1:K]}, Z_{[1:t]}, Y_{[t+1:n]}) \ominus Y_{[1:t]}$.

Proof: The proof follows the same lines of Lemma 10. ■

We consider next some Markov chains that arises naturally when we have 3 interacting nodes and which will be needed for Theorem 6.

Lemma 11 (Markov chains induced by interactive encoding of three nodes): Consider a set of three sources $(X^n, Y^n, Z^n) \sim \prod_{t=1}^n p_{XYZ}(x_t, y_t, z_t)$ and integer $K \in \mathbb{N}$. For each $l \in [1 : K]$ consider arbitrary message sets $\mathcal{I}_x^l, \mathcal{I}_y^l, \mathcal{I}_z^l$ and arbitrary functions

$$f_x^l(X^n, \mathcal{J}_x^{[1:l-1]}, \mathcal{J}_y^{[1:l-1]}, \mathcal{J}_z^{[1:l-1]}) = \mathcal{J}_x^l, \quad (413)$$

$$f_y^l(Y^n, \mathcal{J}_x^{[1:l]}, \mathcal{J}_y^{[1:l-1]}, \mathcal{J}_z^{[1:l-1]}) = \mathcal{J}_y^l, \quad (414)$$

$$f_z^l(Z^n, \mathcal{J}_x^{[1:l]}, \mathcal{J}_y^{[1:l]}, \mathcal{J}_z^{[1:l-1]}) = \mathcal{J}_z^l \quad (415)$$

with $\mathcal{J}_x^l \in \mathcal{I}_x^l, \mathcal{J}_y^l \in \mathcal{I}_y^l$ and $\mathcal{J}_z^l \in \mathcal{I}_z^l$. The following Markov chain relations are valid for each $t \in [1 : n]$ and $l \in [1 : K]$:

- 1) $(\mathcal{J}_x^1, \mathcal{J}_y^1, X_{[1:t-1]}, X_{[t+1:n]}, Y_{[t+1:n]}, Z_{[1:t-1]}) \ominus (X_{[t]}, Y_{[t]}) \ominus Z_{[t]}$,
- 2) $(\mathcal{J}_x^l, \mathcal{J}_y^l, Y_{[1:t-1]}) \ominus (\mathcal{J}_x^{[1:l-1]}, \mathcal{J}_y^{[1:l-1]}, \mathcal{J}_z^{[1:l-1]}, X^n, Y_{[t:n]}, Z_{[1:t-1]}) \ominus Z_{[t]}$,
- 3) $(\mathcal{J}_z^l, Z_{[t+1:n]}) \ominus (\mathcal{J}_x^{[1:l]}, \mathcal{J}_y^{[1:l]}, \mathcal{J}_z^{[1:l-1]}, X^n, Y_{[t+1:n]}, Z_{[1:t]}) \ominus Y_{[t]}$,
- 4) $Z_{[t+1:n]} \ominus (\mathcal{J}_x^{[1:K]}, \mathcal{J}_y^{[1:K]}, \mathcal{J}_z^{[1:K]}, X^n, Y_{[t+1:n]}, Z_{[1:t]}) \ominus Y_{[t]}$,
- 5) $Y_{[1:t-1]} \ominus (\mathcal{J}_x^{[1:K]}, \mathcal{J}_y^{[1:K]}, \mathcal{J}_z^{[1:K]}, X^n, Y_{[t:n]}, Z_{[1:t-1]}) \ominus Z_{[t]}$.

Proof: Along the same lines of Lemma10 and for that reason omitted. ■

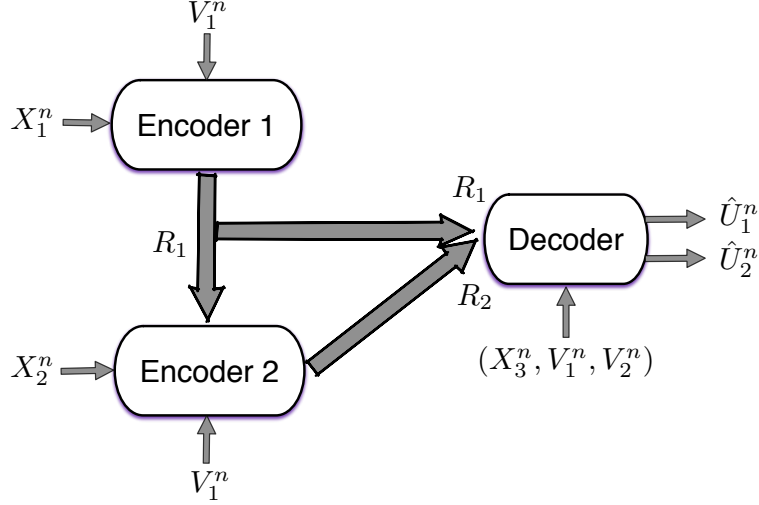


Figure 9: Cooperative Berger-Tung problem.

APPENDIX B

COOPERATIVE BERGER-TUNG PROBLEM WITH SIDE INFORMATION AT THE DECODER

We derive an inner bound on the rate region of the setup described in Fig. 9. It should be emphasized that we will not consider distortion measures, we only focus is on the exchange of descriptions. Encoders 1 and 2 observe source sequences X_1^n and X_2^n , and also have access to a common side information V_1^n . Whereas, the decoder has access to side informations (X_3^n, V_1^n, V_2^n) . Upon observing X_1^n and V_1^n , Encoder 1 generates a message M_1 which is transmitted to Encoder 2 and the decoder. Encoder 2, upon observing (X_2^n, V_1^n) and the message M_1 , generates a message M_2 which is transmitted only to the decoder. Finally, the decoder uses messages (M_1, M_2) and the side informations (X_3^n, V_1^n, V_2^n) to reconstruct two sequences $(\hat{U}_1^n, \hat{U}_2^n)$ which are jointly typical with $(X_1^n, X_2^n, X_3^n, V_1^n, V_2^n)$. In precise terms, we will assume the following:

- A probability mass function $p_{X_1 X_2 X_3 U_1 U_2 V_1 V_2}$ which takes values on cartesian product finite alphabets $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{V}_1 \times \mathcal{V}_2$, and that satisfies the following Markov chains:

$$U_1 \dashv (X_1, V_1) \dashv (X_2, X_3, V_2), \quad U_2 \dashv (U_1, X_2, V_1) \dashv (X_1, X_3, V_2). \quad (416)$$

- Five random vectors $(X_1^n, X_2^n, X_3^n, V_1^n, V_2^n)$ (not necessarily independently and identically distributed with $p_{X_1 X_2 X_3 V_1 V_2}$ which take values on alphabets $\mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{X}_3^n \times \mathcal{V}_1^n \times \mathcal{V}_2^n$

such that, for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr \left\{ (X_1^n, X_2^n, X_3^n, V_1^n, V_2^n) \in \mathcal{T}_{[X_1 X_2 X_3 V_1 V_2] \epsilon}^n \right\} = 1 . \quad (417)$$

Definition 4 (Cooperative code): A code $(n, f_1^n, f_2^n, g^n, \mathcal{M}_1, \mathcal{M}_2)$ for the setup in Fig. 9 is composed by:

- Two set of indices $\mathcal{M}_1, \mathcal{M}_2$.
- An encoding function $f_1^n : \mathcal{X}_1^n \times \mathcal{V}_1^n \rightarrow \mathcal{M}_1$, such that $f_1^n(x_1^n, v_1^n) = m_1$.
- An encoding function $f_2^n : \mathcal{X}_2^n \times \mathcal{V}_1^n \times \mathcal{M}_1 \rightarrow \mathcal{M}_2$, such that $f_2^n(x_2^n, v_1^n, m_1) = m_2$.
- A decoding function $g^n : \mathcal{X}_3^n \times \mathcal{V}_1^n \times \mathcal{V}_2^n \times \mathcal{M}_1 \times \mathcal{M}_2 \rightarrow \mathcal{U}_1^n \times \mathcal{U}_2^n$, such that $g^n(x_3^n, v_1^n, v_2^n, m_1, m_2) = (\hat{u}_1^n, \hat{u}_2^n)$.

Definition 5 (Achievable rates): We say that (R_1, R_2) are ϵ -achievable if there exists a code $(n, f_1^n, f_2^n, g^n, \mathcal{M}_1, \mathcal{M}_2)$ such that:

$$\frac{1}{n} \log \|\mathcal{M}_1\| \leq R_1 + \epsilon , \quad \frac{1}{n} \log \|\mathcal{M}_2\| \leq R_2 + \epsilon \quad (418)$$

and

$$\Pr \left\{ (\hat{U}_1^n, \hat{U}_2^n, V_1^n, V_2^n, X_1^n, X_2^n, X_3^n) \in \mathcal{T}_{[U_1 U_2 V_1 V_2 X_1 X_2 X_3] \epsilon}^n \right\} \leq \epsilon . \quad (419)$$

The closure of the set of all achievable rates (R_1, R_2) is denoted by \mathcal{R}_{CBT} .

The following theorem presents an inner bound to \mathcal{R}_{CBT} .

Theorem 9: (Inner bound on the rate region of the cooperative Berger-Tung problem) Consider $\mathcal{R}_{CBT}^{\text{inner}}$ the closure of the set of rates satisfying:

$$\begin{aligned} R_1 &> I(X_1; U_1 | X_2 V_1) , \\ R_2 &> I(X_2; U_2 | X_3 V_1 V_2 U_1) , \\ R_1 + R_2 &> I(X_1 X_2; U_1 U_2 | X_3 V_1 V_2) , \end{aligned}$$

where the union is over all probability distributions verifying (416). Then $\mathcal{R}_{CBT}^{\text{inner}} \subseteq \mathcal{R}_{CBT}$.

Remark 17: Notice that we are not asking for $(X_1^n, X_2^n, X_3^n, V_1^n, V_2^n)$ to be independently and identically distributed. This is in fact not needed for the result that follows. For us, when trying to use this result, the case of most interest will be when (X_1^n, X_2^n, X_3^n) is generated using the product measure $\prod_{i=1}^n p_{X_1 X_2 X_3}(x_{1i}, x_{2i}, x_{3i})$, (that is, when $(X_1 X_2 X_3)$ is a DMS). However, (V_1^n, V_2^n) will not be independently and identically distributed. Still, (417) will be satisfied.

Remark 18: Notice that unlike the classical rate-distortion problem we are not interested in an average per-symbol distortion constraints at the decoder. We only require that the obtained sequences be jointly typical with the sources. Clearly the problem can be slightly modified to consider the case in which reconstruction distortion constraints are of interest. In fact, case (C) reported in [26], considers a similar setting. Here, given the importance of this result for our interactive scheme, we present a slightly different and more direct proof of the achievability, where we discuss the key points in the encoding and decoding procedures which will be relevant for our extension to the interactive problem.

Proof: Our proof uses standard ideas from multiterminal source coding. As V_1^n is common to both encoders and decoder we can set without loss of generality $V_1^n = \emptyset$. Conditioning with respect to V_1 the final expressions can take into account the situation in which $V_1^n \neq \emptyset$.

A. Coding generation

We randomly generate $2^{n\hat{R}_1}$ codewords $U_1^n(k)$, $k \in [1 : 2^{\hat{R}_1}]$ according to

$$U_1^n(k) \sim \frac{\mathbb{1} \left\{ u_1^n \in \mathcal{T}_{[U_1]^{e_{cd}}}^n \right\}}{\left\| \mathcal{T}_{[U_1]^{e_{cd}}}^n \right\|}, \quad e_{cd} > 0. \quad (420)$$

These $2^{n\hat{R}_1}$ codewords are distributed uniformly over 2^{nR_1} bins denoted by $\mathcal{B}_1(m_1)$, where $m_1 \in [1 : 2^{nR_1}]$. For each codeword $u_1^n(k)$ with $k \in [1 : 2^{n\hat{R}_1}]$, we randomly generate $2^{n\hat{R}_2}$ codewords according to:

$$U_2^n(l, k) \sim \frac{\mathbb{1} \left\{ u_2^n \in \mathcal{T}_{[U_2|U_1]^{e_{cd}}}(u_1^n(k)) \right\}}{\left\| \mathcal{T}_{[U_2|U_1]^{e_{cd}}}(u_1^n(k)) \right\|}, \quad e_{cd} > 0 \quad (421)$$

with $l \in [1 : 2^{n\hat{R}_2}]$. The $2^{n(\hat{R}_1+\hat{R}_2)}$ codewords generated are distributed uniformly in 2^{nR_2} bins, denoted by $\mathcal{B}_2(m_2)$, $m_2 \in [1 : 2^{nR_2}]$. It is worth to mention that the codewords $\{U_2^n(l, k)\}$ are not distributed in a different structure of bins for each k , but on only one *super-bin* structure of size $2^{n(\hat{R}_1+\hat{R}_2)}/2^{nR_2}$ where $\mathcal{B}_2(m_2)$ does not needed to be indexed with k .

As will be clear, this will not constraint the decoder to use *successive decoding* and instead use *joint decoding* in order to recover the desired codewords $(\hat{U}_1^n, \hat{U}_2^n)$. Finally all codebooks are revealed to all parties.

B. Encoding at node 1

Given x_1^n , the encoder search for $k \in [1 : 2^{n\hat{R}_1}]$ in such a way that:

$$(x_1^n, u_1^n(k)) \in \mathcal{T}_{[X_1 U_1] \epsilon_2}^n, \quad \epsilon_2 > 0. \quad (422)$$

If more than one index satisfies this condition, then we choose the one with the smallest index. Otherwise, if no such index exists, we choose an arbitrary one and declare an error. Finally we select m_1 as the index of the bin which contains the codeword $u_1^n(k)$ found and transmit it to nodes 2 and 3.

C. Decoding at node 2

Given x_2^n and m_1 , we search in the bin $\mathcal{B}_1(m_1)$ for an index $k \in [1 : 2^{n\hat{R}_1}]$ such that:

$$(x_2^n, u_1^n(k)) \in \mathcal{T}_{[X_2 U_1] \epsilon_3}^n, \quad \epsilon_3 > 0. \quad (423)$$

If there only one index that satisfies this we declare it as the index generated at node 1. If there several or none we choose a predefined one and declare an error. The chosen index is denoted as $\hat{k}(2)$.

D. Encoding at node 2

Given x_2^n and $\hat{k}(2)$ we search for $l \in [1 : 2^{n\hat{R}_2}]$ such that:

$$(x_2^n, u_1^n(\hat{k}(2)), u_2^n(l, \hat{k}(2))) \in \mathcal{T}_{[X_2 U_1 U_2] \epsilon_4}^n, \quad \epsilon_4 > 0. \quad (424)$$

If more than one index satisfies this condition, then we choose the one with the smallest index. Otherwise, if no such index exists, we choose an arbitrary one and declare an error. Finally we select m_2 as the index of the bin which contains the codeword $u_2^n(l, \hat{k}(2))$ selected and transmit it to node 3.

E. Decoding at node 3

Given x_3^n , v_2^n and m_1, m_2 , the decoder search in the bins $\mathcal{B}_1(m_1)$ and $\mathcal{B}_2(m_2)$ for a pair of indices $(k, l) \in [1 : 2^{n\hat{R}_1}] \times [1 : 2^{n\hat{R}_2}]$ such that

$$(x_3^n, v_2^n, u_1^n(k), u_2^n(l, k)) \in \mathcal{T}_{[X_3 V_2 U_1 U_2] \epsilon}^n, \quad \epsilon > 0. \quad (425)$$

If there only one pair of indices that satisfy this we declare it as the indices generated at node 1 and 2. If there several or none we choose a predefined pair and declare an error. The chosen pair is denoted by $(\hat{k}(3), \hat{l}(3))$. Finally, the decoder declares $(\hat{u}_1^n, \hat{u}_2^n) = (u_1^n(\hat{k}(3)), u_2^n(\hat{l}(3), \hat{k}(3)))$.

F. Error probability analysis

Consider (K, L) the description indices generated at node 1 and 2, and (M_1, M_2) the corresponding bin indices. With $\hat{K}(2)$ and $(\hat{K}(3), \hat{L}(3))$ we denote the indices recovered at nodes 2 and 3. We want to prove that $\Pr\{\mathcal{E}\} \leq \epsilon'$ when n is sufficiently large, where

$$\mathcal{E} = \left\{ \left(X_1^n, X_2^n, X_3^n, V_2^n, U_1^n(\hat{K}(3)), U_2^n(\hat{L}(3), \hat{K}(3)) \right) \notin \mathcal{T}_{[X_1 X_2 X_3 V_2 U_1 U_2]^\epsilon}^n \right\}. \quad (426)$$

We consider the following events of error:

- $\mathcal{E}_1 = \left\{ (X_1^n, X_2^n, X_3^n, V_2^n) \notin \mathcal{T}_{[X_1 X_2 X_3 V_2]^{\epsilon_1}}^n \right\}, \epsilon_1 > 0.$
- $\mathcal{E}_2 = \left\{ (X_1^n, U_1^n(k)) \notin \mathcal{T}_{[X_1 U_1]^{\epsilon_2}}^n \forall k \in [1 : 2^{n\hat{R}_1}] \right\}, \epsilon_2 > 0.$
- $\mathcal{E}_3 = \left\{ (X_1^n, X_2^n, X_3^n, V_2^n, U_1^n(K)) \notin \mathcal{T}_{[X_1 X_2 X_3 V_2 U_1]^{\epsilon_3}}^n \right\}, \epsilon_3 > 0.$
- $\mathcal{E}_4 = \left\{ \exists \hat{k} \neq K, \hat{k} \in \mathcal{B}_1(M_1), (X_2^n, U_1^n(\hat{k})) \in \mathcal{T}_{[X_1 U_1]^{\epsilon_3}}^n \right\}, \epsilon_3 > 0.$
- $\mathcal{E}_5 = \left\{ (X_2^n, U_1^n(\hat{K}(2)), U_2^n(l, \hat{K}(2))) \notin \mathcal{T}_{[X_2 U_1 U_2]^{\epsilon_4}}^n \forall l \in [1 : 2^{n\hat{R}_2}] \right\}, \epsilon_4 > 0.$
- $\mathcal{E}_6 = \left\{ (X_1^n, X_2^n, X_3^n, V_2^n, U_1^n(K), U_2^n(L, \hat{K}(2))) \notin \mathcal{T}_{[X_1 X_2 X_3 V_2 U_1 U_2]^\epsilon}^n \right\}, \epsilon > 0.$
- $\mathcal{E}_7 = \left\{ \exists \hat{k} \neq K, \hat{l} \neq L, \hat{k} \in \mathcal{B}_1(M_1), \hat{l} \in \mathcal{B}_2(M_2), (X_3^n, V_2^n, U_1^n(\hat{k}), U_2^n(\hat{l}, \hat{k})) \in \mathcal{T}_{[X_3 V_2 U_1 U_2]^\epsilon}^n \right\}.$

Clearly $\mathcal{E} \subseteq \bigcup_{i=1}^7 \mathcal{E}_i$. In fact, it is easy to show that $\left\{ (\hat{K}(3), \hat{L}(3)) \neq (K, L), \hat{K}(2) \neq K \right\} \subseteq \bigcup_{i=1}^7 \mathcal{E}_i$. From hypothesis, we obtain that $\lim_{n \rightarrow \infty} \Pr\{\mathcal{E}_1\} = 0$. Choosing $\epsilon_1 < \frac{\epsilon_2}{\|\mathcal{U}_1\|}$ and $\epsilon_2 < \epsilon_{cd}$ we can use Lemma 5 and its Corollary (with the following equivalences: $V \equiv U_1, X \equiv X_1, U \equiv \emptyset$) to obtain $\lim_{n \rightarrow \infty} \Pr\{\mathcal{E}_2\} = 0$ if

$$\hat{R}_1 > I(U_1; X_1) + \delta(\epsilon_1, \epsilon_2, \epsilon_{cd}, n). \quad (427)$$

For the analysis of $\Pr\{\mathcal{E}_3\}$ we can use Lemma 7, its Corollary and Lemma 8 defining $Y \equiv X_2 X_3$, $X \equiv X_1$ and $U \equiv U_1$ and using ϵ_2, ϵ_3 and ϵ_{cd} sufficiently small¹³ to obtain $\lim_{n \rightarrow \infty} \Pr\{\mathcal{E}_3\} = 0$.

For the analysis of $\Pr\{\mathcal{E}_4\}$ we can write:

$$\begin{aligned} \Pr\{\mathcal{E}_4\} &= \mathbb{E}[\Pr\{\mathcal{E}_4 | K = k, M_1 = m_1\}] \\ &= \mathbb{E} \left[\Pr \left\{ \bigcup_{\substack{\hat{k} \neq k \\ \hat{k} \in \mathcal{B}_1(m_1)}} \left\{ (X_2^n, U_1^n(\hat{k})) \in \mathcal{T}_{[X_1 U_1]^{\epsilon_3}}^n \right\} \middle| \begin{matrix} K=k \\ M_1=m_1 \end{matrix} \right\} \right]. \end{aligned} \quad (428)$$

¹³In the following, we will not indicate anymore the corresponding values of the constants ϵ , the arguments of δ and the equivalence between the involved random variables in order to use the lemmas from Appendix A

Using Lemma 6 (with the appropriate equivalences on the involved random variables) and the statistical properties of the codebooks, binning and encoding, we have, that for each k, m_1 :

$$\lim_{n \rightarrow \infty} \Pr \left\{ \bigcup_{\hat{k} \neq k, \hat{k} \in \mathcal{B}_1(m_1)} \left\{ (X_2^n, U_1^n(\hat{k})) \in \mathcal{T}_{[X_1 U_1] \epsilon_3}^n \right\} \middle| \begin{matrix} K=k \\ M_1=m_1 \end{matrix} \right\} = 0 \quad (429)$$

provided that

$$\frac{1}{n} \log \mathbb{E} \|\mathcal{B}_1(m_1)\| < I(X_2; U_1) - \delta(\epsilon_1, \epsilon_3, \epsilon_{cd}, n) . \quad (430)$$

As $\mathbb{E} [\|\mathcal{B}_1(m_1)\|] = 2^{n(\hat{R}_1 - R_1)} \forall m_1$ we have that $\lim_{n \rightarrow \infty} \Pr \{\mathcal{E}_4\} = 0$ provided that

$$\hat{R}_1 - R_1 < I(X_2; U_1) - \delta . \quad (431)$$

The analysis of $\Pr \{\mathcal{E}_5\}$ follows the same lines of $\Pr \{\mathcal{E}_2\}$. The above analysis implies that

$$\lim_{n \rightarrow \infty} \Pr \left\{ (X_2^n, U_1^n(\hat{K}(2))) \in \mathcal{T}_{[X_2 U_1] \epsilon_3}^n \right\} = 1 . \quad (432)$$

Then, by Lemma 5 and its Corollary we have that $\lim_{n \rightarrow \infty} \Pr \{\mathcal{E}_5\} = 0$ if :

$$\hat{R}_2 > I(X_2; U_2|U_1) + \delta . \quad (433)$$

From Lemmas 7 and 8, similarly as with $\Pr \{\mathcal{E}_3\}$, we have $\lim_{n \rightarrow \infty} \Pr \{\mathcal{E}_6\} = 0$. Let us turn to analyze $\Pr \{\mathcal{E}_7\}$:

$$\begin{aligned} \Pr \{\mathcal{E}_7\} &= \mathbb{E} [\Pr \{\mathcal{E}_7 | K = k, L = l, M_1 = m_1, M_2 = m_2\}] \\ &= \mathbb{E} \left[\Pr \left\{ \bigcup_{\substack{(\hat{k}, \hat{l}) \neq (k, l) \\ \hat{k} \in \mathcal{B}_1(m_1) \\ (\hat{k}, \hat{l}) \in \mathcal{B}_2(m_2)}} \left\{ (X_3^n, V_2^n, U_1^n(\hat{k}), U_2^n(\hat{l}, \hat{k})) \in \mathcal{T}_{[X_3 V_2 U_1 U_2] \epsilon}^n \right\} \middle| \begin{matrix} K=k, L=l \\ M_1=m_1, M_2=m_2 \end{matrix} \right\} \right] \\ &\leq \mathbb{E} [\alpha_1 + \alpha_2 + \alpha_3] \end{aligned} \quad (434)$$

where we have

$$\alpha_1 = \Pr \left\{ \bigcup_{\substack{\hat{k} \neq k \\ \hat{k} \in \mathcal{B}_1(m_1) \\ (\hat{k}, l) \in \mathcal{B}_2(m_2)}} \left\{ (X_3^n, V_2^n, U_1^n(\hat{k}), U_2^n(l, \hat{k})) \in \mathcal{T}_{[X_3 V_2 U_1 U_2] \epsilon}^n \right\} \middle| \begin{matrix} K=k, L=l \\ M_1=m_1, M_2=m_2 \end{matrix} \right\} , \quad (435)$$

$$\alpha_2 = \Pr \left\{ \bigcup_{\substack{\hat{l} \neq l \\ (k, \hat{l}) \in \mathcal{B}_2(m_2)}} \left\{ \left(X_3^n, V_2^n, U_1^n(k), U_2^n(\hat{l}, k) \right) \in \mathcal{T}_{[X_3 V_2 U_1 U_2]^\epsilon}^n \right\} \middle| \begin{matrix} K=k, L=l \\ M_1=m_1, M_2=m_2 \end{matrix} \right\} , \quad (436)$$

$$\alpha_3 = \Pr \left\{ \bigcup_{\substack{\hat{k} \neq k, \hat{l} \neq l \\ \hat{k} \in \mathcal{B}_1(m_1) \\ (\hat{k}, \hat{l}) \in \mathcal{B}_2(m_2)}} \left\{ \left(X_3^n, V_2^n, U_1^n(\hat{k}), U_2^n(\hat{l}, \hat{k}) \right) \in \mathcal{T}_{[X_3 V_2 U_1 U_2]^\epsilon}^n \right\} \middle| \begin{matrix} K=k, L=l \\ M_1=m_1, M_2=m_2 \end{matrix} \right\} . \quad (437)$$

We can use the Lemma 6 to obtain:

$$\lim_{n \rightarrow \infty} \alpha_1 = 0 , \quad \lim_{n \rightarrow \infty} \alpha_2 = 0 , \quad \lim_{n \rightarrow \infty} \alpha_3 = 0 \quad (438)$$

provided that

$$\frac{1}{n} \log \mathbb{E} \left\| \hat{k} : (\hat{k}, l) \in \mathcal{B}_2(m_2), \hat{k} \in \mathcal{B}_1(m_1) \right\| < I(X_3 V_2; U_1 U_2) - \delta, \quad (439)$$

$$\frac{1}{n} \log \mathbb{E} \left\| \hat{l} : (k, \hat{l}) \in \mathcal{B}_2(m_2) \right\| < I(X_3 V_2; U_2 | U_1) - \delta , \quad (440)$$

$$\frac{1}{n} \log \mathbb{E} \left\| (\hat{k}, \hat{l}) : (\hat{k}, \hat{l}) \in \mathcal{B}_2(m_2), \hat{k} \in \mathcal{B}_1(m_1) \right\| < I(X_3 V_2; U_1 U_2) - \delta . \quad (441)$$

Because on how the binning is performed, we have:

$$\mathbb{E} \left\| \hat{k} : (\hat{k}, l) \in \mathcal{B}_2(m_2), \hat{k} \in \mathcal{B}_1(m_1) \right\| = 2^{n(\hat{R}_1 - R_1 - R_2)} , \quad (442)$$

$$\mathbb{E} \left\| \hat{l} : (k, \hat{l}) \in \mathcal{B}_2(m_2) \right\| = 2^{n(\hat{R}_2 - R_2)} , \quad (443)$$

$$\mathbb{E} \left\| (\hat{k}, \hat{l}) : (\hat{k}, \hat{l}) \in \mathcal{B}_2(m_2), \hat{k} \in \mathcal{B}_1(m_1) \right\| = 2^{n(\hat{R}_1 + \hat{R}_2 - R_1 - R_2)} , \quad (444)$$

which give us:

$$(\hat{R}_1 - R_1) - R_2 < I(X_3 V_2; U_1 U_2) - \delta , \quad (445)$$

$$\hat{R}_2 - R_2 < I(X_3 V_2; U_2 | U_1) - \delta , \quad (446)$$

$$(\hat{R}_1 + \hat{R}_2) - (R_1 + R_2) < I(X_3 V_2; U_1 U_2) - \delta . \quad (447)$$

Notice that equation (445) remains inactive because of (447). Equations (427), (431), (433), (446), (447) can be combined with:

$$\hat{R}_1 > R_1 , \quad (448)$$

$$\hat{R}_1 + \hat{R}_2 > R_2 , \quad (449)$$

which follow from the binning structure assumed in the generated codebooks. A Fourier-Motzkin elimination procedure can be done to eliminate \hat{R}_1 and \hat{R}_2 obtaining the desired rate region (conditioning also the mutual information terms on V_1). ■

The following Corollary considers the case in which a genie gives node 2 the value of M_1 . Indeed, this case will be important for our main result.

Corollary 5: If a genie gives M_1 to node 2, the achievable region $\mathcal{R}_{CBT}^{\text{inner}}$ reduces to:

$$R_2 > I(X_2; U_2 | X_3 V_1 V_2 U_1) , \quad (450)$$

$$R_1 + R_2 > I(X_1 X_2; U_1 U_2 | X_3 V_1 V_2) . \quad (451)$$

The proof of this result is straightforward and thus it will not be presented.

APPENDIX C

PROOF OF THEOREM 1

Let us describe the coding generation, encoding and decoding procedures. We will consider the following notation. With $M_{i \rightarrow S, l}$ we will denote the index corresponding to the true description $U_{i \rightarrow S, l}^n$ generated at node i at round l and destined to the group of nodes $S \in \mathcal{C}(\mathcal{M})$ with $i \notin S$. With $\hat{M}_{i \rightarrow S, l}(j)$ where $S \in \mathcal{C}(\mathcal{M})$, $i \notin S$, $j \in S$ we denote the corresponding estimated index at node j .

A. Codebook generation

Consider the round $l \in [1 : K]$. For simplicity let us consider the descriptions at node 1. We generate $2^{n\hat{R}_{1 \rightarrow 23}^{(l)}}$ independent and identically distributed n -length codewords $U_{1 \rightarrow 23, l}^n(m_{1 \rightarrow 23, l}, m_{\mathcal{W}_{[1, l]}})$ according to:

$$U_{1 \rightarrow 23, l}^n(m_{1 \rightarrow 23, l}, m_{\mathcal{W}_{[1, l]}}) \sim \frac{\mathbb{1} \left\{ u_{1 \rightarrow 23, l}^n \in \mathcal{T}_{[U_{1 \rightarrow 23, l}^n | \mathcal{W}_{[1, l]}}^n \right\}_{\epsilon(1, 23, l)} \left(w_{[1, l]}^n \right) \right\}}{\left\| \mathcal{T}_{[U_{1 \rightarrow 23, l}^n | \mathcal{W}_{[1, l]}}^n \right\}_{\epsilon(1, 23, l)} \left(w_{[1, l]}^n \right) \right\|}, \quad \epsilon(1, 23, l) > 0 \quad (452)$$

where $m_{1 \rightarrow 23, l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 23}^{(l)}}]$ and let $m_{\mathcal{W}_{[1, l]}}$ denote the indices of the common descriptions $\mathcal{W}_{[1, l]}^n$ generated in rounds $t \in [1 : l - 1]$. For example, $m_{\mathcal{W}_{[1, l]}} = \{m_{1 \rightarrow 23, t}, m_{2 \rightarrow 13, t}, m_{3 \rightarrow 12, t}\}_{t=1}^{l-1}$. With $w_{[1, l]}^n$ we denote the set of n -length common information codewords from previous rounds corresponding to the indices $m_{\mathcal{W}_{[1, l]}}$. For each $m_{\mathcal{W}_{[3, l-1]}}$ consider the set of $2^{n(\hat{R}_{1 \rightarrow 23}^{(l)} + \hat{R}_{3 \rightarrow 12}^{(l)})}$ codewords $U_{1 \rightarrow 23, l}^n(m_{1 \rightarrow 23, l}, m_{3 \rightarrow 12, l-1}, m_{\mathcal{W}_{[3, l-1]}})$. These n -length codewords are distributed independently

and uniformly over $2^{n\hat{R}_{1 \rightarrow 23}^{(l)}}$ bins denoted by $\mathcal{B}_{1 \rightarrow 23,l} \left(p_{1 \rightarrow 23,l}, m_{\mathcal{W}_{[3,l-1]}} \right)$ with $p_{1 \rightarrow 23,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 23}^{(l)}}]$. Notice that this binning structure is exactly the same we used for the cooperative Berger-Tung problem in Appendix B. Node 1 distributes codewords $U_{1 \rightarrow 23,l}^n(m_{1 \rightarrow 23,l}, m_{3 \rightarrow 12,l-1}, m_{\mathcal{W}_{[3,l-1]}})$ in a super-binning structure. This will allow node 2 to recover both, $m_{1 \rightarrow 23,l}$ and $m_{3 \rightarrow 12,l-1}$, using the same procedure as in the Berger-Tung problem described above. Notice that a different super-binning structure is generated for every $m_{\mathcal{W}_{[3,l-1]}}$. This is without loss of generality, because at round l nodes 1, 2 and 3, will have a very good estimated of it (see below).

We also generate $2^{n\hat{R}_{1 \rightarrow 2}^{(l)}}$ and $2^{n\hat{R}_{1 \rightarrow 3}^{(l)}}$ independent and identically distributed n -length codewords $U_{1 \rightarrow 2,l}^n(m_{1 \rightarrow 2,l}, m_{\mathcal{W}_{[2,l]}})$, and $U_{1 \rightarrow 3,l}^n(m_{1 \rightarrow 3,l}, m_{\mathcal{W}_{[2,l]}})$ according to:

$$U_{1 \rightarrow 2,l}^n(m_{1 \rightarrow 2,l}, m_{\mathcal{W}_{[2,l]}}) \sim \frac{\mathbb{1} \left\{ u_{1 \rightarrow 2,l}^n \in \mathcal{T}_{[U_{1 \rightarrow 2,l}^n | \mathcal{W}_{[2,l]} \mathcal{V}_{[12,l,1]}] \epsilon(1,2,l)}^n \left(w_{[2,l]}^n, v_{[12,l,1]}^n \right) \right\}}{\left\| \mathcal{T}_{[U_{1 \rightarrow 2,l}^n | \mathcal{W}_{[2,l]} \mathcal{V}_{[12,l,1]}] \epsilon(1,2,l)}^n \left(w_{[2,l]}^n, v_{[12,l,1]}^n \right) \right\|}}, \quad (453)$$

$$U_{1 \rightarrow 3,l}^n(m_{1 \rightarrow 3,l}, m_{\mathcal{W}_{[2,l]}}) \sim \frac{\mathbb{1} \left\{ u_{1 \rightarrow 3,l}^n \in \mathcal{T}_{[U_{1 \rightarrow 3,l}^n | \mathcal{W}_{[2,l]} \mathcal{V}_{[13,l,1]}] \epsilon(1,3,l)}^n \left(w_{[2,l]}^n, v_{[13,l,1]}^n \right) \right\}}{\left\| \mathcal{T}_{[U_{1 \rightarrow 3,l}^n | \mathcal{W}_{[2,l]} \mathcal{V}_{[13,l,1]}] \epsilon(1,3,l)}^n \left(w_{[2,l]}^n, v_{[13,l,1]}^n \right) \right\|}}, \quad (454)$$

where $\epsilon(1, 2, l) > 0$, $\epsilon(1, 3, l) > 0$, and $m_{1 \rightarrow 2,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 2}^{(l)}}]$ and $m_{1 \rightarrow 3,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 3}^{(l)}}]$. These codewords are distributed uniformly on $2^{n\hat{R}_{1 \rightarrow 2}^{(l)}}$ bins denoted by $\mathcal{B}_{1 \rightarrow 2,l}(p_{1 \rightarrow 2,l}, m_{\mathcal{W}_{[2,l]}})$ and indexed with $p_{1 \rightarrow 2,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 2}^{(l)}}]$ and on $2^{n\hat{R}_{1 \rightarrow 3}^{(l)}}$ bins denoted by $\mathcal{B}_{1 \rightarrow 3,l}(p_{1 \rightarrow 3,l}, m_{\mathcal{W}_{[2,l]}})$ and indexed with $p_{1 \rightarrow 3,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 3}^{(l)}}]$, respectively. Notice that these codewords (which will be used to generated private descriptions to node 2 and 3) are not distributed in a super-binning structure. This is because there is not explicit cooperation between the nodes at this level. That is, node 2 is not compelled to recover the private description that node 1 generate for node 3, and for that reason the private description that node 2 generate for node 3 is not superimposed over the former. Notice that the binning structure used for the codewords to be utilized by node 1 impose the following relationships:

$$R_{1 \rightarrow 23}^{(l)} < \hat{R}_{1 \rightarrow 23}^{(l)} + \hat{R}_{3 \rightarrow 12}^{(l-1)}, \quad (455)$$

$$R_{1 \rightarrow 2}^{(l)} < \hat{R}_{1 \rightarrow 2}^{(l)}, \quad (456)$$

$$R_{1 \rightarrow 3}^{(l)} < \hat{R}_{1 \rightarrow 3}^{(l)}. \quad (457)$$

The common and private codewords to be utilized in nodes 2 and 3, for every round, are generated by following a similar procedure and theirs corresponding rates have analogous relationships. After this is finished the generated codebooks are revealed to all the nodes in the network.

B. Encoding technique

Consider node 1 at round $l \in [1 : K]$. Upon observing x_1^n and given all of its encoding and decoding history up to round l , encoder 1 first looks for a codeword $u_{1 \rightarrow 23,l}^n(m_{1 \rightarrow 23,l}, \hat{m}_{\mathcal{W}_{[1,l]}}(1))$ such that $\epsilon_c(1, 23, l) > 0$,

$$\left(x_1^n, w_{[1,l]}^n(\hat{m}_{\mathcal{W}_{[1,l]}}(1)), u_{1 \rightarrow 23,l}^n(m_{1 \rightarrow 23,l}, \hat{m}_{\mathcal{W}_{[1,l]}}(1)) \right) \in \mathcal{T}_{[U_{1 \rightarrow 23,l} X_1 \mathcal{W}_{[1,l]}] \epsilon_c(1,23,l)}^n. \quad (458)$$

Notice that some components in $\hat{m}_{\mathcal{W}_{[1,l]}}(1)$ are generated at node 1 and are perfectly known. If more than one codeword satisfies this condition, then we choose the one with the smallest index. Otherwise, if no such codeword exists, we choose an arbitrary index and declare an error. With the chosen index $m_{1 \rightarrow 23,l}$, and with $\hat{m}_{\mathcal{W}_{[3,l-1]}}(1)$, we determine the index $p_{1 \rightarrow 23,l}$ of the bin $\mathcal{B}_{1 \rightarrow 23,l}(p_{1 \rightarrow 23,l}, \hat{m}_{\mathcal{W}_{[3,l-1]}}(1))$ to which $u_{1 \rightarrow 23,l}^n(m_{1 \rightarrow 23,l}, \hat{m}_{\mathcal{W}_{[3,l-1]}}(1), \hat{m}_{\mathcal{W}_{[3,l-1]}}(1))$ belongs. After this, Encoder 1 generates the private descriptions looking for codewords $u_{1 \rightarrow 2,l}^n(m_{1 \rightarrow 2,l}, \hat{m}_{\mathcal{W}_{[2,l]}}(1), \hat{m}_{\mathcal{V}_{[12,l,1]}}(1)), u_{1 \rightarrow 3,l}^n(m_{1 \rightarrow 2,l}, \hat{m}_{\mathcal{W}_{[2,l]}}(1), \hat{m}_{\mathcal{V}_{[13,l,1]}}(1))$ such that

$$\begin{aligned} & \left(x_1^n, w_{[2,l]}^n(\hat{m}_{\mathcal{W}_{[2,l]}}(1)), v_{[12,l,1]}^n(\hat{m}_{\mathcal{V}_{[12,l,1]}}(1)), u_{1 \rightarrow 2,l}^n(m_{1 \rightarrow 2,l}, \hat{m}_{\mathcal{W}_{[2,l]}}(1), \hat{m}_{\mathcal{V}_{[12,l,1]}}(1)) \right) \\ & \in \mathcal{T}_{[U_{1 \rightarrow 2,l} X_1 \mathcal{W}_{[2,l]} \mathcal{V}_{[12,l,1]}] \epsilon_c(1,2,l)}^n, \end{aligned} \quad (459)$$

$$\begin{aligned} & \left(x_1^n, w_{[2,l]}^n(\hat{m}_{\mathcal{W}_{[2,l]}}(1)), v_{[13,l,1]}^n(\hat{m}_{\mathcal{V}_{[13,l,1]}}(1)), u_{1 \rightarrow 3,l}^n(m_{1 \rightarrow 2,l}, \hat{m}_{\mathcal{W}_{[2,l]}}(1), \hat{m}_{\mathcal{V}_{[13,l,1]}}(1)) \right) \\ & \in \mathcal{T}_{[U_{1 \rightarrow 3,l} X_1 \mathcal{W}_{[2,l]} \mathcal{V}_{[13,l,1]}] \epsilon_c(1,3,l)}^n, \end{aligned} \quad (460)$$

respectively, where $\epsilon_c(1, 2, l) > 0$ and $\epsilon_c(1, 3, l) > 0$. Given $(\hat{m}_{\mathcal{W}_{[2,l]}}(1), \hat{m}_{\mathcal{V}_{[12,l,1]}}(1), \hat{m}_{\mathcal{V}_{[13,l,1]}}(1))$, the encoding procedure continues by determining the bin indices $p_{1 \rightarrow 2,l}$ and $p_{1 \rightarrow 3,l}$ to which the generated private descriptions belong to. Node 1 then transmits to node 2 and 3 the indices $(p_{1 \rightarrow 23,l}, p_{1 \rightarrow 2,l}, p_{1 \rightarrow 3,l})$. The encoding in nodes 2 and 3 follows along the same lines and for that reason are not described.

C. Decoding technique

Consider round $l \in [1 : K + 1]$ and node 2. During round the present and previous round node 2 receives $(p_{1 \rightarrow 23,l}, p_{3 \rightarrow 12,l-1}, p_{1 \rightarrow 2,l}, p_{1 \rightarrow 3,l}, p_{3 \rightarrow 1,l-1}, p_{3 \rightarrow 2,l-1})$. However, only the indices $(p_{1 \rightarrow 23,l}, p_{3 \rightarrow 12,l-1}, p_{1 \rightarrow 2,l}, p_{3 \rightarrow 2,l-1})$ are the ones relevant to him. Knowing this set of indices, node 2 aims to recover the exact values of $(m_{1 \rightarrow 23,l}, m_{3 \rightarrow 12,l-1}, m_{1 \rightarrow 2,l}, m_{3 \rightarrow 2,l-1})$. This is done through

successive decoding where first, the common information indices are recovered by looking for the unique tuple of codewords $u_{1 \rightarrow 23,l}^n(m_{1 \rightarrow 23,l}, m_{3 \rightarrow 12,l-1}, \hat{m}_{\mathcal{W}_{[3,l-1]}}(2))$, $u_{3 \rightarrow 12,l-1}^n(m_{3 \rightarrow 12,l-1}, \hat{m}_{\mathcal{W}_{[3,l-1]}}(2))$ that satisfies:

$$\begin{aligned} & \left(x_2^n, w_{[3,l-1]}^n(\hat{m}_{\mathcal{W}_{[3,l-1]}}(2)), v_{[12,l,1]}^n(\hat{m}_{\mathcal{V}_{[12,l,1]}}(2)), v_{[23,l-1,3]}^n(\hat{m}_{\mathcal{V}_{[23,l-1,3]}}(2)), \right. \\ & \left. u_{1 \rightarrow 23,l}^n(m_{1 \rightarrow 23,l}, m_{3 \rightarrow 12,l-1}, \hat{m}_{\mathcal{W}_{[3,l-1]}}(2)), u_{3 \rightarrow 12,l-1}^n(m_{3 \rightarrow 12,l-1}, \hat{m}_{\mathcal{W}_{[3,l-1]}}(2)) \right) \\ & \in \mathcal{T}_{[U_{1 \rightarrow 23,l} U_{3 \rightarrow 12,l-1} X_2 \mathcal{W}_{[3,l-1]} \mathcal{V}_{[23,l-1,3]} \mathcal{V}_{[12,l,1]}] \epsilon_{dc}(2,l)}, \quad \epsilon_{dc}(2,l) > 0 \end{aligned} \quad (461)$$

and also belong to the bins indicated by $p_{1 \rightarrow 23,l}$ and $p_{3 \rightarrow 12,l-1}$. If there are more than one pair of codewords, or none that satisfies this, we choose a predefined one and declare an error. After this is done, node 2 can recover the private information indices by looking at codewords $u_{1 \rightarrow 2,l}^n(m_{1 \rightarrow 2,l}, \hat{m}_{\mathcal{W}_{[2,l]}}(2), \hat{m}_{\mathcal{V}_{[12,l,1]}}(2))$ and $u_{3 \rightarrow 2,l-1}^n(m_{3 \rightarrow 2,l-1}, \hat{m}_{\mathcal{W}_{[1,l]}}(2), \hat{m}_{\mathcal{V}_{[23,l-1,3]}}(2))$ which satisfy

$$\begin{aligned} & \left(x_2^n, w_{[2,l]}^n(\hat{m}_{\mathcal{W}_{[2,l]}}(2)), v_{[12,l,1]}^n(\hat{m}_{\mathcal{V}_{[12,l,1]}}(2)), v_{[23,l-1,3]}^n(\hat{m}_{\mathcal{V}_{[23,l-1,3]}}(2)), \right. \\ & \left. u_{1 \rightarrow 2,l}^n(m_{1 \rightarrow 2,l}, \hat{m}_{\mathcal{W}_{[2,l]}}(2), \hat{m}_{\mathcal{V}_{[12,l,1]}}(2)), u_{3 \rightarrow 2,l-1}^n(m_{3 \rightarrow 2,l-1}, \hat{m}_{\mathcal{W}_{[1,l]}}(2), \hat{m}_{\mathcal{V}_{[23,l-1,3]}}(2)) \right) \\ & \in \mathcal{T}_{[U_{1 \rightarrow 2,l} U_{3 \rightarrow 2,l-1} X_2 \mathcal{W}_{[2,l]} \mathcal{V}_{[23,l-1,3]} \mathcal{V}_{[12,l,1]}] \epsilon_{dp}(2,l)}, \quad \epsilon_{dp}(2,l) > 0 \end{aligned} \quad (462)$$

and are in the bins given by $p_{1 \rightarrow 2,l}$ and $p_{3 \rightarrow 2,l-1}$. If there are more than one pair of codewords, or none that satisfies this, we choose a predefined one and declare an error. The decoding in nodes 1 and 3 is exactly the same and for that reason are not described.

D. Lossy reconstructions

When the exchange of information is completed, each node needs to estimate the other nodes sources. For instance, node 1 reconstruct the source of node 2 by computing:

$$\hat{x}_{12,i} = g_{12}(x_{1i}, v_{[12,K+1,1]i}, w_{[1,K+1]i}), \quad i = 1, 2, \dots, n, \quad (463)$$

and similarly, for the source of node 3:

$$\hat{x}_{13,i} = g_{13}(x_{1i}, v_{[13,K+1,1]i}, w_{[1,K+1]i}), \quad i = 1, 2, \dots, n. \quad (464)$$

Reconstruction at nodes 2 and 3 is done in a similar way using the adequate reconstruction functions.

E. Error and distortion analysis

In order to maintain expressions simple, in the following when we denote a description without the corresponding index, i.e. $U_{i \rightarrow S, l}^n$ or $\mathcal{W}_{[1, l]}^n$, we will assume that the corresponding index is the true one generated in the corresponding nodes through the detailed encoding procedure. Consider round l and the event $\mathcal{D}_l = \mathcal{G}_l \cap \mathcal{F}_l$, where for $\epsilon_l > 0$,

$$\mathcal{G}_l = \left\{ (X_1^n, X_2^n, X_3^n, \mathcal{W}_{[1, l]}^n, \mathcal{V}_{[12, l, 1]}^n, \mathcal{V}_{[13, l, 1]}^n, \mathcal{V}_{[23, l, 2]}^n) \in \mathcal{T}_{[X_1 X_2 X_3 \mathcal{W}_{[1, l]} \mathcal{V}_{[12, l, 1]} \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]}] \epsilon_l}^n \right\}, \quad (465)$$

$$\mathcal{F}_l = \left\{ \hat{M}_{i \rightarrow S, t}(j) = M_{i \rightarrow S, t}, \quad S \in \mathcal{C}(\mathcal{M}), \quad i \notin S, \quad j \in S, \quad t \in [1 : l - 1], \text{ with exception of } \right. \\ \left. \hat{M}_{3 \rightarrow 12, l-1}(2), \hat{M}_{3 \rightarrow 2, l-1}(2) \right\}. \quad (466)$$

The set \mathcal{G}_l indicates that all the descriptions generated in the network, up to round l are jointly typical with the sources. The occurrence of this depends mainly on the encoding procedure in the nodes. Set \mathcal{F}_l indicates, that up to round l , all nodes were able to recover the true indices of the descriptions. This clearly implies that there were not errors at the decoding procedures in all the nodes in the network. The condition in \mathcal{F}_l on $\hat{M}_{3 \rightarrow 12, l-1}(2)$, $\hat{M}_{3 \rightarrow 2, l-1}(2)$ is due to the fact, that the decoding of those descriptions in node 2 occurs during round l . The occurrence of \mathcal{D}_l guarantees that at the beginning of round l :

- Node 1 and 2 share a common path of descriptions $\mathcal{W}_{[1, l]}^n \cup \mathcal{V}_{[12, l, 1]}^n$ which are typical with (X_1^n, X_2^n, X_3^n) .
- Node 1 and 3 share a common path of descriptions $\mathcal{W}_{[1, l]}^n \cup \mathcal{V}_{[13, l, 1]}^n$ which are typical with (X_1^n, X_2^n, X_3^n) .
- Node 2 and 3 share a common path of descriptions $\mathcal{W}_{[3, l-1]}^n \cup \mathcal{V}_{[23, l-1, 3]}^n$ which are typical with (X_1^n, X_2^n, X_3^n) .

Let us also define the event \mathcal{E}_l :

$$\mathcal{E}_l = \{ \text{there exists at least an error at the encoding or decoding in a node during round } l \} \\ = \bigcup_{i \in \mathcal{M}} \mathcal{E}_{enc}(i, l) \cup \mathcal{E}_{dec}(i, l) \quad (467)$$

where $\mathcal{E}_{enc}(i, l)$ contains the errors at the encoding in node i during round l and $\mathcal{E}_{dec}(i, l)$ considers the event that at node i during round l there is a failure at recovering an index generated previously in other node. For example, at node 1 and during round l :

$$\mathcal{E}_{enc}(i, l) = \mathcal{E}_{enc}(1, l, 23) \cup \mathcal{E}_{enc}(1, l, 2) \cup \mathcal{E}_{enc}(1, l, 3) \quad (468)$$

where

$$\mathcal{E}_{enc}(1, l, 23) = \left\{ \left(X_1^n, \mathcal{W}_{[1,l]}^n(\hat{M}_{\mathcal{W}_{[1,l]}}(1)) U_{1 \rightarrow 23,l}^n(m_{1 \rightarrow 23,l}, \hat{M}_{\mathcal{W}_{[1,l]}}(1)) \right) \notin \mathcal{T}_{[U_{1 \rightarrow 23,l} X_1 \mathcal{W}_{[1,l]}] \epsilon_c(1,l,23)}^n \right. \\ \left. \forall m_{1 \rightarrow 23,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 23}^{(l)}}] \right\} \quad (469)$$

$$\mathcal{E}_{enc}(1, l, 2) = \left\{ \left(X_1^n, \mathcal{W}_{[2,l]}^n(\hat{M}_{\mathcal{W}_{[2,l]}}(1)), \mathcal{V}_{[12,l,1]}^n(\hat{M}_{\mathcal{V}_{[12,l,1]}}(1)), U_{1 \rightarrow 2,l}^n(m_{1 \rightarrow 2,l}, \hat{M}_{\mathcal{W}_{[2,l]}}(1), \hat{M}_{\mathcal{V}_{[12,l,1]}}(1)) \right) \right. \\ \left. \notin \mathcal{T}_{[U_{1 \rightarrow 2,l} X_1 \mathcal{W}_{[2,l]} \mathcal{V}_{[12,l,1]}] \epsilon_c(1,l,2)}^n \forall m_{1 \rightarrow 2,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 2}^{(l)}}] \right\} \quad (470)$$

$$\mathcal{E}_{enc}(1, l, 3) = \left\{ \left(X_1^n, \mathcal{W}_{[2,l]}^n(\hat{M}_{\mathcal{W}_{[2,l]}}(1)), \mathcal{V}_{[13,l,1]}^n(\hat{M}_{\mathcal{V}_{[13,l,1]}}(1)), U_{1 \rightarrow 3,l}^n(m_{1 \rightarrow 3,l}, \hat{M}_{\mathcal{W}_{[2,l]}}(1), \hat{M}_{\mathcal{V}_{[13,l,1]}}(1)) \right) \right. \\ \left. \notin \mathcal{T}_{[U_{1 \rightarrow 3,l} X_1 \mathcal{W}_{[2,l]} \mathcal{V}_{[13,l,1]}] \epsilon_c(1,l,3)}^n \forall m_{1 \rightarrow 3,l} \in [1 : 2^{n\hat{R}_{1 \rightarrow 3}^{(l)}}] \right\} . \quad (471)$$

Event $\mathcal{E}_{dec}(i, l)$ can be decomposed as:

$$\mathcal{E}_{dec}(i, l) = \bigcup_{S \in \mathcal{C}(\mathcal{M}), i \in S} \bigcup_{j: j \notin S} \left\{ \hat{M}_{j \rightarrow S,l}(i) \neq M_{j \rightarrow S,l} \right\} . \quad (472)$$

At the end of the information exchange phase we would expect the occurrence of $\mathcal{D}_{K+1} \cap \bar{\mathcal{E}}_{K+1}$, where \mathcal{E}_{K+1} is the event of an error during round $K+1$. As during round $K+1$ only node 2 tries to recover the descriptions generated during round K in node 3, we have:

$$\mathcal{E}_{K+1} = \mathcal{E}_{dec}(2, K+1) = \left\{ \hat{M}_{3 \rightarrow 12,K}(2) \neq M_{3 \rightarrow 12,K} \text{ or } \hat{M}_{3 \rightarrow 2,K}(2) \neq M_{3 \rightarrow 2,K} \right\} . \quad (473)$$

The occurrence of $\mathcal{D}_{K+1} \cap \bar{\mathcal{E}}_{K+1}$ guarantees that all the descriptions generated during the K rounds of information exchange in the network are jointly typical with the sources realizations and that those descriptions can be perfectly recovered in all the nodes. In this way, if we can guarantee that $\Pr \{ \mathcal{D}_{K+1} \cap \bar{\mathcal{E}}_{K+1} \} \xrightarrow[n \rightarrow \infty]{} 1$, then with probability converging to one we obtain:

- Node 1 and 2 share a common path of descriptions $\mathcal{W}_{[1,K+1]}^n \cup \mathcal{V}_{[12,K+1,1]}^n$ which are typical with (X_1^n, X_2^n, X_3^n) .
- Node 1 and 3 share a common path of descriptions $\mathcal{W}_{[1,K+1]}^n \cup \mathcal{V}_{[13,K+1,1]}^n$ which are typical with (X_1^n, X_2^n, X_3^n) .
- Node 2 and 3 share a common path of descriptions $\mathcal{W}_{[1,K+1]}^n \cup \mathcal{V}_{[23,K+1,2]}^n$ which are typical with (X_1^n, X_2^n, X_3^n) .

Using standard analysis ideas, the average distortions (over the codebooks) at the reconstruction stages in all the nodes satisfy the required fidelity constraints. From there is straightforward to prove the existence of good codebooks for the network. In order to prove that

$\Pr \{ \mathcal{D}_{K+1} \cap \bar{\mathcal{E}}_{K+1} \} \xrightarrow{n \rightarrow \infty} 1$ let us write:

$$\begin{aligned}
\Pr \{ \overline{\mathcal{D}_{K+1} \cap \bar{\mathcal{E}}_{K+1}} \} &= \Pr \{ \bar{\mathcal{D}}_{K+1} \cup \mathcal{E}_{K+1} \} = \Pr \{ \bar{\mathcal{D}}_{K+1} \} + \Pr \{ \mathcal{D}_{K+1} \cap \mathcal{E}_{K+1} \} \\
&\leq \Pr \{ \bar{\mathcal{D}}_{K+1} \cap \mathcal{D}_K \} + \Pr \{ \bar{\mathcal{D}}_K \} + \Pr \{ \mathcal{D}_{K+1} \cap \mathcal{E}_{K+1} \} \\
&\leq \Pr \{ \bar{\mathcal{D}}_K \} + \Pr \{ \bar{\mathcal{D}}_{K+1} \cap (\mathcal{D}_K \cap \bar{\mathcal{E}}_K) \} + \Pr \{ \mathcal{D}_K \cap \mathcal{E}_K \} + \Pr \{ \mathcal{D}_{K+1} \cap \mathcal{E}_{K+1} \} \\
&\leq \Pr \{ \bar{\mathcal{D}}_1 \} + \sum_{l=1}^{K+1} \Pr \{ \mathcal{D}_l \cap \mathcal{E}_l \} + \sum_{l=1}^K \Pr \{ \bar{\mathcal{D}}_{l+1} \cap (\mathcal{D}_l \cap \bar{\mathcal{E}}_l) \} . \tag{474}
\end{aligned}$$

Notice that

$$\mathcal{D}_1 = \{ (X_1^n, X_2^n, X_3^n) \in \mathcal{T}_{[X_1 X_2 X_3] \epsilon_1}^n \} , \quad \epsilon_1 > 0 . \tag{475}$$

From Lemma 2, we see that for every $\epsilon_1 > 0$ $\Pr \{ \bar{\mathcal{D}}_1 \} \xrightarrow{n \rightarrow \infty} 0$. Then, it is easy to see that $\Pr \{ \mathcal{D}_{K+1} \cap \bar{\mathcal{E}}_{K+1} \} \xrightarrow{n \rightarrow \infty} 1$ will hold if the coding generation, the encoding and decoding procedures described above allow us to have the following:

- 1) If $\Pr \{ \mathcal{D}_l \} \xrightarrow{n \rightarrow \infty} 1$ then $\Pr \{ \mathcal{D}_{l+1} \} \xrightarrow{n \rightarrow \infty} 1 \quad \forall l \in [1 : K + 1]$.
- 2) $\Pr \{ \mathcal{D}_l \cap \mathcal{E}_l \} \xrightarrow{n \rightarrow \infty} 0 \quad \forall l \in [1 : K + 1]$.

In the following we will prove these facts. Observe that, at round l the nodes act sequentially:

Encoding at node 1 \rightarrow Decoding at node 2 $\rightarrow \dots \rightarrow$ Encoding at node 3 \rightarrow Decoding at node 1.

Then, using (467) we can write condition 2) as:

$$\begin{aligned}
\Pr \{ \mathcal{D}_l \cap \mathcal{E}_l \} &= \Pr \{ \mathcal{D}_l \cap \mathcal{E}_{enc}(1, l) \} + \Pr \{ \mathcal{D}_l \cap \mathcal{E}_{dec}(2, l) \cap \bar{\mathcal{E}}_{enc}(1, l) \} \\
&\quad + \Pr \{ \mathcal{D}_l \cap \mathcal{E}_{enc}(2, l) \cap \bar{\mathcal{E}}_{enc}(1, l) \cap \bar{\mathcal{E}}_{dec}(2, l) \} + \\
&\quad \dots + \Pr \{ \mathcal{D}_l \cap \mathcal{E}_{dec}(1, l) \cap \bar{\mathcal{E}}_{enc}(1, l) \cap \dots \cap \bar{\mathcal{E}}_{enc}(3, l) \} . \tag{476}
\end{aligned}$$

Assume then that at the beginning of round l we have $\Pr \{ \mathcal{D}_l \} \xrightarrow{n \rightarrow \infty} 1$. Let us analyze the encoding procedure at node 1. Let us consider $\Pr \{ \mathcal{D}_l \cap \mathcal{E}_{enc}(1, l) \}$. We can write:

$$\begin{aligned}
\Pr \{ \mathcal{D}_l \cap \mathcal{E}_{enc}(1, l) \} &\leq \Pr \{ \mathcal{E}_{enc}(1, l, 23) \cap \mathcal{D}_l \} + \Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{D}_l \cap \bar{\mathcal{E}}_{enc}(1, l, 23) \} \\
&\quad + \Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \bar{\mathcal{E}}_{enc}(1, l, 23) \} . \tag{477}
\end{aligned}$$

From Lemma 1 and the fact that $\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}_l \} = 1$ we have that $\lim_{n \rightarrow \infty} \Pr \{ \mathcal{A}_l(1, 23) \} = 1$ where:

$$\mathcal{A}_l(1, 23) = \left\{ (X_1^n, \mathcal{W}_{[1, l]}^n) \in \mathcal{T}_{[X_1 \mathcal{W}_{[1, l]}] \epsilon_l}^n \right\} . \tag{478}$$

Then, we can use Lemma 5 to obtain:

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{E}_{enc}(1, l, 23) \cap \mathcal{D}_l \} = 0 \quad (479)$$

provided that

$$\hat{R}_{1 \rightarrow 23}^{(l)} > I \left(X_1; U_{1 \rightarrow 23, l} \middle| \mathcal{W}_{[1, l]} \right) + \delta_c(1, l, 23) \quad (480)$$

where $\delta_c(1, l, 23)$ can be made arbitrarily small. On the other hand we can write:

$$\Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{D}_l \cap \bar{\mathcal{E}}_{enc}(1, l, 23) \} \leq \Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{G}_l(1, 2) \cap \mathcal{F}_l \} + \Pr \{ \bar{\mathcal{G}}_l(1, 2) \} \quad (481)$$

where

$$\mathcal{G}_l(1, 2) = \left\{ (X_1^n, X_2^n, X_3^n, \mathcal{W}_{[2, l]}^n, \mathcal{V}_{[12, l, 1]}^n, \mathcal{V}_{[13, l, 1]}^n, \mathcal{V}_{[23, l, 2]}^n) \in \mathcal{T}_{[X_1 X_2 X_3 \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 1]} \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]}] \epsilon_l(1, 2)}^n \right\} \quad (482)$$

where $\epsilon_l(1, 2) > 0$. As explained before, $\Pr \{ \mathcal{G}_l \} \xrightarrow{n \rightarrow \infty} 1$. Then, from condition (480) we have

$$\Pr \{ \bar{\mathcal{E}}_{enc}(1, l, 23) \cap \mathcal{F}_l \} = \Pr \left\{ (X_1^n, \mathcal{W}_{[2, l]}^n) \in \mathcal{T}_{[X_1 \mathcal{W}_{[2, l]}] \epsilon_c(1, l, 23)} \right\} \xrightarrow{n \rightarrow \infty} 1. \quad (483)$$

Moreover, from the coding generation and the encoding procedure proposed is immediate to use Lemma 8 to show that:

$$\begin{aligned} & \Pr (U_{1 \rightarrow 23, l}^n = u_{1 \rightarrow 23, l}^n | x_1^n, w_{[1, l]}^n, \bar{\mathcal{E}}_{enc}(1, l, 23) \cap \mathcal{F}_l) = \\ & \frac{\mathbb{1} \left\{ u_{1 \rightarrow 23, l}^n \in \mathcal{T}_{[U_{1 \rightarrow 23, l} | X_1 \mathcal{W}_{[1, l]}] \epsilon_c(1, 23, l)}^n (x_1^n, w_{[1, l]}^n) \right\}}{\left\| \mathcal{T}_{[U_{1 \rightarrow 23, l} | X_1 \mathcal{W}_{[1, l]}] \epsilon_c(1, 23, l)}^n (x_1^n, w_{[1, l]}^n) \right\|}. \end{aligned} \quad (484)$$

Then, from Markov chain

$$U_{1 \rightarrow 23, l} \ominus (X_1, \mathcal{W}_{[1, l]}) \ominus (X_2, X_3, \mathcal{V}_{[12, l, 1]}, \mathcal{V}_{[13, l, 1]}, \mathcal{V}_{[23, l, 2]}) \quad (485)$$

and the Markov Lemma 7, for sufficiently small $(\epsilon_c(1, l, 23), \epsilon_l, \epsilon_l(1, 2))$ and after some minor manipulations, we can obtain:

$$\Pr \{ \mathcal{G}_l(1, 2) \} \xrightarrow{n \rightarrow \infty} 1. \quad (486)$$

From equation (481) it is clear that we need to analyze term $\Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{G}_l(1, 2) \cap \mathcal{F}_l \}$.

Similarly as before $\lim_{n \rightarrow \infty} \Pr \{ \mathcal{A}_l(1, 2) \}$ where:

$$\mathcal{A}_l(1, 2) = \left\{ (X_1^n, \mathcal{W}_{[2, l]}^n, \mathcal{V}_{[12, l, 1]}^n) \in \mathcal{T}_{[X_1 \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 1]}] \epsilon_l(1, 2)}^n \right\}, \quad (487)$$

which allow us to write:

$$\Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{G}_l(1, 2) \cap \mathcal{F}_l \} \leq \Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{A}_l(1, 2) \cap \mathcal{F}_l \} . \quad (488)$$

Using again Lemma 5, we obtain that $\Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{G}_l(1, 2) \cap \mathcal{F}_l \} \xrightarrow[n \rightarrow \infty]{} 0$ provided that

$$\hat{R}_{1 \rightarrow 2}^{(l)} > I \left(X_1; U_{1 \rightarrow 2, l} \middle| \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 1]} \right) + \delta_c(1, l, 2) \quad (489)$$

where $\delta_c(1, l, 2)$ can be made arbitrarily small. For the analysis of $\Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \bar{\mathcal{E}}_{enc}(1, l, 23) \}$ we follow the same procedure. We can write

$$\Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \bar{\mathcal{E}}_{enc}(1, l, 23) \} \leq \Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{G}_l(1, 3) \cap \mathcal{F}_l \} + \Pr \{ \bar{\mathcal{G}}_l(1, 3) \} \quad (490)$$

with

$$\mathcal{G}_l(1, 3) = \left\{ (X_1^n, X_2^n, X_3^n, \mathcal{W}_{[2, l]}^n, \mathcal{V}_{[12, l, 2]}^n, \mathcal{V}_{[13, l, 1]}^n, \mathcal{V}_{[23, l, 2]}^n) \in \mathcal{T}_{[X_1 X_2 X_3 \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]}] \epsilon_l(1, 3)}^n \right\} . \quad (491)$$

Using the Markov chain

$$U_{1 \rightarrow 2, l} \ominus (X_1, \mathcal{W}_{[2, l]}, \mathcal{V}_{[12, l, 1]}) \ominus (X_2, X_3, \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]}) , \quad (492)$$

the fact that $\Pr \{ \mathcal{G}_l(1, 2) \} \xrightarrow[n \rightarrow \infty]{} 1$ and the Markov Lemma 7, and Lemma 8 for appropriately chosen values of $(\epsilon_c(1, l, 2), \epsilon_l(1, 2), \epsilon_l(1, 3))$ we have:

$$\Pr \{ \mathcal{G}_l(1, 3) \} \xrightarrow[n \rightarrow \infty]{} 1 . \quad (493)$$

Following exactly the same reasoning as above, we have that in order to have

$$\Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \bar{\mathcal{E}}_{enc}(1, l, 23) \} \xrightarrow[n \rightarrow \infty]{} 0 , \quad (494)$$

besides conditions (480) and (489) we need:

$$\hat{R}_{1 \rightarrow 3}^{(l)} > I \left(X_1; U_{1 \rightarrow 3, l} \middle| \mathcal{W}_{[2, l]} \mathcal{V}_{[13, l, 1]} \right) + \delta_c(1, l, 3) \quad (495)$$

for sufficiently small $\delta_c(1, l, 3)$. With these conditions we have proved that the encoding procedure in node 1 during round l permit us to have:

$$\Pr \{ \mathcal{D}_l \cap \mathcal{E}_{enc}(1, l) \} \xrightarrow[n \rightarrow \infty]{} 0 . \quad (496)$$

Another instance of the Markov lemma, jointly with Markov chain

$$U_{1 \rightarrow 3, l} \ominus (X_1, \mathcal{W}_{[2, l]}, \mathcal{V}_{[13, l, 1]}) \ominus (X_2, X_3, \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[23, l, 2]}) \quad (497)$$

and Lemma 8 allow us to have:

$$\Pr \{ \mathcal{G}_l(2, 13) \} \xrightarrow[n \rightarrow \infty]{} 1 \quad (498)$$

where

$$\mathcal{G}_l(2, 13) = \left\{ (X_1^n, X_2^n, X_3^n, \mathcal{W}_{[2,l]}^n, \mathcal{V}_{[12,l,2]}^n, \mathcal{V}_{[13,l,3]}^n, \mathcal{V}_{[23,l,2]}^n) \in \mathcal{T}_{[X_1 X_2 X_3 \mathcal{W}_{[2,l]} \mathcal{V}_{[12,l,2]} \mathcal{V}_{[13,l,3]} \mathcal{V}_{[23,l,2]}] \epsilon_l(2,13)}^n \right\} . \quad (499)$$

At this point we have to analyze the decoding in node 2. If that decoding is successful, with (498), the analysis of the encoding at node 2 follows the same lines as above¹⁴. The same can be said of the encoding at node 3 (after successful decoding). In this way, we terminate round l with

$$\Pr \{ \mathcal{D}_{l+1} \} = \Pr \{ \mathcal{G}_{l+1} \cap \mathcal{F}_{l+1} \} \xrightarrow[n \rightarrow \infty]{} 1 \quad (500)$$

which is one of the results we wanted. Clearly, analyzing now the decoding at node 2 (from which we can easily extrapolate the analysis to the decoding at node 1 and 3) we will be able to obtain $\Pr \{ \mathcal{D}_l \cap \mathcal{E}_l \} \xrightarrow[n \rightarrow \infty]{} 0$ which is the other required result.

The decoding in each of nodes follows the approach of *successive decoding*. Decoder 2 will try to find first the common descriptions $M_{1 \rightarrow 23,l}$ and $M_{3 \rightarrow 12,l-1}$. Then, it will try to find the private descriptions $M_{1 \rightarrow 2,l}$ and $M_{3 \rightarrow 2,l-1}$ (using of course the previously obtained common descriptions as side information). Clearly, the use of *joint-decoding* could improve the rate region. However, the analysis of this strategy, besides of being more difficult to analyze, it will give rise to more complex rate region. It can be easily seen, that the joint-decoding region will contain several sum-rate equations that will contain common and private rates. Successive decoding allows for a rate region where the sum-rate equations contain solely common rates or private rates, being more easy to analyze and understand.

In order to analyze the decoding, we can write:

$$\Pr \{ \mathcal{D}_l \cap \mathcal{E}_{dec}(2, l) \cap \bar{\mathcal{E}}_{enc}(1, l) \} \leq \Pr \{ \mathcal{E}_{dec}(2, l) \cap \mathcal{F}_l \cap \mathcal{G}_l(2, 13) \} + \Pr \{ \bar{\mathcal{G}}_l(2, 13) \} . \quad (501)$$

As $\Pr \{ \mathcal{G}_l(2, 13) \} \xrightarrow[n \rightarrow \infty]{} 1$ we can concentrate our effort on the first term. Event $\mathcal{E}_{dec}(2, l)$ can be written as:

$$\mathcal{E}_{dec}(2, l) = \mathcal{H}_{common}(2, l) \cup \mathcal{H}_{private}(2, l) , \quad (502)$$

¹⁴See that $\mathcal{G}_l(2, 13)$ has, for the encoding at node 2 the same role that \mathcal{G}_l has for the encoding at node 1 during round l .

where

$$\mathcal{H}_{common}(2, l) = \left\{ \left(\hat{M}_{3 \rightarrow 12, l-1}(2), \hat{M}_{1 \rightarrow 23, l}(2) \right) \neq (M_{3 \rightarrow 12, l-1}, M_{1 \rightarrow 23, l}) \right\}, \quad (503)$$

$$\mathcal{H}_{private}(2, l) = \left\{ \left(\hat{M}_{3 \rightarrow 2, l-1}(2), \hat{M}_{1 \rightarrow 2, l}(2) \right) \neq (M_{3 \rightarrow 2, l-1}, M_{1 \rightarrow 2, l}) \right\}. \quad (504)$$

From these definitions, we can easily deduce that:

$$\begin{aligned} \Pr \{ \mathcal{E}_{dec}(2, l) \cap \mathcal{F}_l \cap \mathcal{G}_l(2, 13) \} &= \Pr \{ \mathcal{H}_{common}(2, l) \cap \mathcal{F}_l \cap \mathcal{G}_l(2, 13) \} \\ &\quad + \Pr \{ \mathcal{H}_{private}(2, l) \cap \mathcal{F}_l \cap \mathcal{G}_l(2, 13) \cap \bar{\mathcal{H}}_{common}(2, l) \} \\ &\leq \Pr \{ \mathcal{K}_{common}(2, l) \} + \Pr \{ \mathcal{K}_{private}(2, l) \} \end{aligned} \quad (505)$$

where

$$\begin{aligned} \mathcal{K}_{common}(2, l) &= \left\{ \exists (\tilde{m}_{1 \rightarrow 23, l}, \tilde{m}_{3 \rightarrow 12, l-1}) \neq (M_{1 \rightarrow 23, l}, M_{3 \rightarrow 12, l-1}), (\tilde{m}_{1 \rightarrow 23, l}, \tilde{m}_{3 \rightarrow 12, l-1}) \right. \\ &\quad \in \mathcal{B}_{1 \rightarrow 23, l}(P_{1 \rightarrow 23, l}) \times \mathcal{B}_{3 \rightarrow 12, l-1}(P_{3 \rightarrow 12, l-1}) : (X_2^n, \mathcal{W}_{[3, l-1]}^n, \mathcal{V}_{[23, l-1, 3]}^n, \mathcal{V}_{[12, l, 1]}^n, \\ &\quad \left. U_{1 \rightarrow 23, l}^n(\tilde{m}_{1 \rightarrow 23, l}, \tilde{m}_{3 \rightarrow 12, l-1}), U_{3 \rightarrow 12, l-1}^n(\tilde{m}_{3 \rightarrow 12, l-1}) \in \mathcal{T}_{[X_1 \mathcal{W}_{[2, l]}^n \mathcal{V}_{[12, l, 1]}^n \mathcal{V}_{[23, l-1, 3]}^n] \epsilon_{dc}(2, l)} \right\}, \end{aligned} \quad (506)$$

$$\begin{aligned} \mathcal{K}_{private}(2, l) &= \left\{ \exists (\tilde{m}_{1 \rightarrow 2, l}, \tilde{m}_{3 \rightarrow 2, l-1}) \neq (M_{1 \rightarrow 2, l}, M_{3 \rightarrow 2, l-1}), (\tilde{m}_{1 \rightarrow 2, l}, \tilde{m}_{3 \rightarrow 2, l-1}) \right. \\ &\quad \in \mathcal{B}_{1 \rightarrow 2, l}(P_{1 \rightarrow 2, l}) \times \mathcal{B}_{3 \rightarrow 2, l-1}(P_{3 \rightarrow 2, l-1}) : (X_2^n, \mathcal{W}_{[2, l]}^n, \mathcal{V}_{[23, l-1, 3]}^n, \mathcal{V}_{[12, l, 1]}^n, \\ &\quad \left. U_{1 \rightarrow 2, l}^n(\tilde{m}_{1 \rightarrow 2, l}), U_{3 \rightarrow 2, l-1}^n(\tilde{m}_{3 \rightarrow 2, l-1}) \in \mathcal{T}_{[X_1 \mathcal{W}_{[2, l]}^n \mathcal{V}_{[12, l, 2]}^n \mathcal{V}_{[23, l, 2]}^n] \epsilon_{dp}(2, l)} \right\}, \end{aligned} \quad (507)$$

where $\epsilon_{dc}(2, l)$, $\epsilon_{dp}(2, l)$ are carefully chosen¹⁵, and for a saving of notation we considered only the indices to be recovered, i.e., $U_{1 \rightarrow 23, l}^n(\tilde{m}_{1 \rightarrow 23, l}, \tilde{m}_{3 \rightarrow 12, l-1}) \equiv U_{1 \rightarrow 23, l}^n(\tilde{m}_{1 \rightarrow 23, l}, \tilde{m}_{3 \rightarrow 12, l-1}, M_{\mathcal{W}_{[3, l-1]}})$. Consider first the recovering of the common information. Node 2 has to recover two indices from a binning structure as the one in the cooperative Berger-Tung problem described in Appendix B.

In Fig. 10, we have a representation of the problem seen at decoder 2. Node 3 generate a common description at rate $R_{3 \rightarrow 12}^{(l-1)}$ using $\mathcal{W}_{[3, l-1]}^n$ as side information. Similarly node 1,

¹⁵ Using Lemma 1 to have:

$$\begin{aligned} \mathcal{G}_l(2, 13) &\subseteq \left(X_2^n, \mathcal{W}_{[2, l]}^n, \mathcal{V}_{[23, l-1, 3]}^n, \mathcal{V}_{[12, l, 1]}^n \right) \in \mathcal{T}_{[X_1 \mathcal{W}_{[2, l]}^n \mathcal{V}_{[12, l, 1]}^n \mathcal{V}_{[23, l-1, 3]}^n] \epsilon_{dc}(2, l)} \\ \mathcal{G}_l(2, 13) &\subseteq \left(X_2^n, \mathcal{W}_{[2, l]}^n, \mathcal{V}_{[23, l, 2]}^n, \mathcal{V}_{[12, l, 2]}^n \right) \in \mathcal{T}_{[X_1 \mathcal{W}_{[2, l]}^n \mathcal{V}_{[12, l, 2]}^n \mathcal{V}_{[23, l, 2]}^n] \epsilon_{dp}(2, l)}. \end{aligned}$$

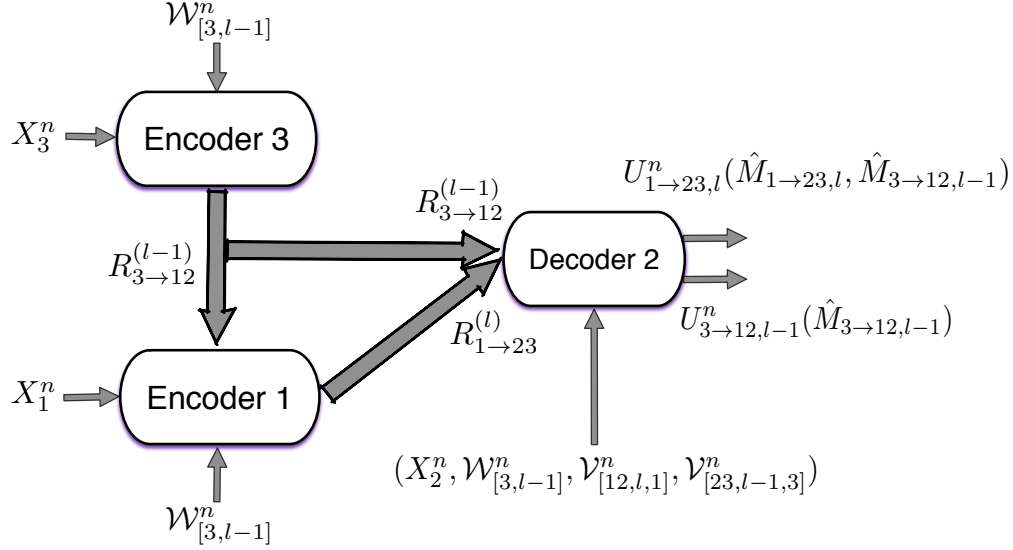


Figure 10: Cooperative Berger-Tung decoding problem for node 2.

after decoding the common description from node 3, generate its own description using the recovered one and also $\mathcal{W}_{[3,l-1]}^n$ as side information. All these operations all done using the super-binning structure as in the cooperative Berger-Tung problem in Appendix B. Then, node 2, using $(X_3^n, \mathcal{W}_{[3,l-1]}^n, \mathcal{V}_{[12,l,1]}^n, \mathcal{V}_{[23,l-1,3]}^n)$ as side information tries to recover the descriptions generated at node 3 and 1. Remember the fact that the encoding procedure at nodes 1 and 3 requires:

$$\hat{R}_{3 \rightarrow 12}^{(l-1)} > I(X_3; U_{3 \rightarrow 12, l-1} | \mathcal{W}_{[3, l-1]}^n) + \delta_c(1, l-1, 12) , \quad (508)$$

$$\hat{R}_{1 \rightarrow 23}^{(l)} > I(X_1; U_{1 \rightarrow 23, l} | \mathcal{W}_{[1, l]}^n) + \delta_c(1, l, 23) , \quad (509)$$

$$R_{1 \rightarrow 23}^{(l)} < \hat{R}_{1 \rightarrow 23}^{(l)} + \hat{R}_{3 \rightarrow 12}^{(l-1)} \quad (510)$$

and that the following Markov chains:

$$U_{3 \rightarrow 12, l-1} \ominus (X_3, \mathcal{W}_{[3, l-1]}^n) \ominus (X_1, X_2, \mathcal{V}_{[12, l, 1]}^n, \mathcal{V}_{[23, l-1, 3]}^n) , \quad (511)$$

$$U_{1 \rightarrow 23, l} \ominus (X_1, \mathcal{W}_{[1, l]}^n) \ominus (X_2, X_3, \mathcal{V}_{[12, l, 1]}^n, \mathcal{V}_{[23, l, 2]}^n) / , \quad (512)$$

are implied by the Markov chains in the conditions of Theorem 1. In this way, we can use the results in Appendix B to show that the following rates imply $\Pr \{ \mathcal{K}_{common}(2, l) \} \xrightarrow[n \rightarrow \infty]{} 0$:

$$R_{1 \rightarrow 23}^{(l)} > I(X_1; U_{1 \rightarrow 23, l} | X_2 \mathcal{W}_{[1, l]}^n \mathcal{V}_{[23, l-1, 3]}^n \mathcal{V}_{[12, l, 1]}^n) + \delta_{dc}(2, l) , \quad (513)$$

$$R_{1 \rightarrow 23}^{(l)} + R_{3 \rightarrow 12}^{(l-1)} > I(X_1 X_3; U_{1 \rightarrow 23, l} U_{3 \rightarrow 12, l-1} | X_2 \mathcal{W}_{[3, l-1]}^n \mathcal{V}_{[23, l-1, 3]}^n \mathcal{V}_{[12, l, 1]}^n) + \delta'_{dc}(2, l) , \quad (514)$$

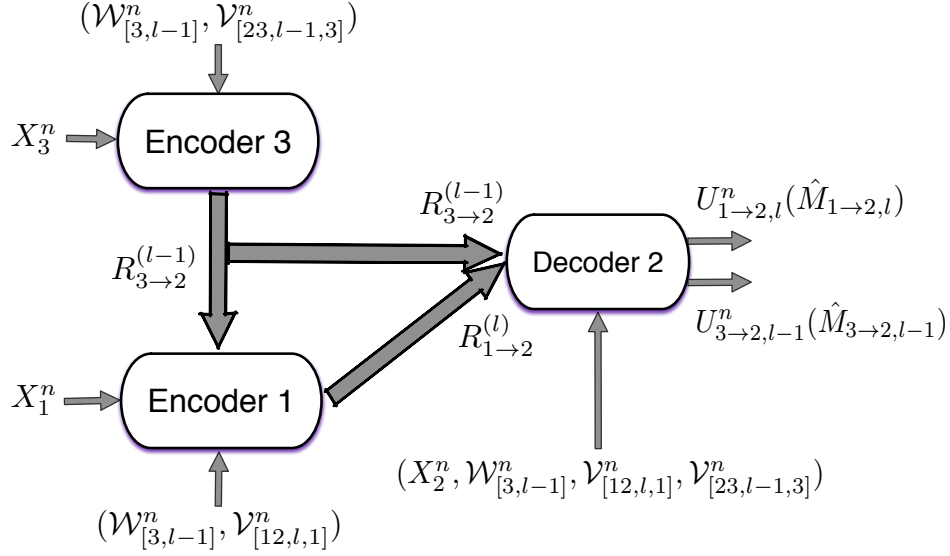


Figure 11: Berger-Tung decoding problem for node 2 when it tries to recover the private descriptions generated in nodes 1 and 3.

where $\delta_{dc}(2, l), \delta'_{dc}(2, l)$ can be made arbitrarily small¹⁶.

The decoding of the private descriptions can be seen as a standard Berger-Tung decoding problem (see Fig. 11) where the binning used to transmit the descriptions generated in node 3 and 1 is not cooperative (in the sense of Theorem 9) as in the case of the common descriptions. Lemma 6 can be easily used to analyze $\Pr \{\mathcal{K}_{private}(2, l)\}$. The following conditions guarantee that $\Pr \{\mathcal{K}_{private}(2, l)\} \xrightarrow{n \rightarrow \infty} 0$:

¹⁶Here we considered the Corollary to Theorem 9. That is we assumed, that node 1 knows perfectly the value of $M_{3 \rightarrow 12, l-1}$. This follows from the assumed fact, that at the beginning of round l , the probability of decoding errors at previous rounds in all nodes goes to zero when $n \rightarrow \infty$. In this way, the constraint on rate $R_{3 \rightarrow 12, l-1}$ that should be considered, according to Theorem 9 is not needed. In fact, constraints on rate $R_{3 \rightarrow 12, l-1}$ will arise when at node 1 we consider the recovering of $M_{3 \rightarrow 12, l-1}$ and $M_{2 \rightarrow 13, l-1}$. For that reason, the analysis carried on is valid. Through this analysis we avoid carrying a lengthy and difficult Fourier-Motzkin procedure to eliminate $\hat{R}_{1 \rightarrow 23}^{(l)}, \hat{R}_{2 \rightarrow 13}^{(l)}, \hat{R}_{3 \rightarrow 12}^{(l)}$ for $l = [1 : K]$.

$$\hat{R}_{1 \rightarrow 2}^{(l)} < R_{1 \rightarrow 2}^{(l)} + I \left(U_{1 \rightarrow 2, l}; X_2 \mathcal{V}_{[23, l, 2]} \middle| \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 1]} \right) - \delta_{dp}(2, l) \quad (515)$$

$$\hat{R}_{3 \rightarrow 2}^{(l-1)} < R_{3 \rightarrow 2}^{(l-1)} + I \left(U_{3 \rightarrow 2, l-1}; X_2 U_{1 \rightarrow 23, l} \mathcal{V}_{[12, l, 2]} \middle| \mathcal{W}_{[1, l]} \mathcal{V}_{[23, l-1, 3]} \right) - \delta'_{dp}(2, l) \quad (516)$$

$$\begin{aligned} \hat{R}_{3 \rightarrow 2}^{(l-1)} + \hat{R}_{1 \rightarrow 2}^{(l)} &< R_{3 \rightarrow 2}^{(l-1)} + R_{1 \rightarrow 2}^{(l)} + I \left(U_{1 \rightarrow 2, l}; X_2 \mathcal{V}_{[23, l, 2]} \middle| \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 1]} \right) \\ &+ I \left(U_{3 \rightarrow 2, l-1}; X_2 U_{1 \rightarrow 23, l} \mathcal{V}_{[12, l, 2]} \middle| \mathcal{W}_{[1, l]} \mathcal{V}_{[23, l-1, 3]} \right) \\ &- I \left(U_{3 \rightarrow 2, l-1}; U_{1 \rightarrow 2, l} \middle| \mathcal{W}_{[2, l]} \mathcal{V}_{[23, l-1, 3]} \mathcal{V}_{[12, l, 1]} X_2 \right) - \delta''_{dp}(2, l) \end{aligned} \quad (517)$$

where $\delta_{dp}(2, l)$, $\delta'_{dp}(2, l)$, $\delta''_{dp}(2, l)$ can be made arbitrarily small. Then, combining all the obtained results, we have that:

$$\Pr \{ \mathcal{D}_l \cap \mathcal{E}_{dec}(2, l) \cap \bar{\mathcal{E}}_{enc}(1, l) \} \xrightarrow{n \rightarrow \infty} 0. \quad (518)$$

At this point, the story is as it was at the encoding stage in node 1 and all the steps can be repeated with minor modifications, proving the desired results at the end of round l :

$$\Pr \{ \mathcal{D}_{l+1} \} \xrightarrow{n \rightarrow \infty} 1, \quad \Pr \{ \mathcal{D}_l \cap \mathcal{E}_l \} \xrightarrow{n \rightarrow \infty} 0. \quad (519)$$

The other rates equations are as follows:

- Encoding at node 2:

$$\hat{R}_{2 \rightarrow 13}^{(l)} > I \left(X_2; U_{2 \rightarrow 13, l} \middle| \mathcal{W}_{[2, l]} \right) + \delta_c(2, l, 13) \quad (520)$$

$$\hat{R}_{2 \rightarrow 1}^{(l)} > I \left(X_2; U_{2 \rightarrow 1, l} \middle| \mathcal{W}_{[3, l]} \mathcal{V}_{[12, l, 2]} \right) + \delta_c(2, l, 1) \quad (521)$$

$$\hat{R}_{2 \rightarrow 3}^{(l)} > I \left(X_2; U_{2 \rightarrow 3, l} \middle| \mathcal{W}_{[3, l]} \mathcal{V}_{[23, l, 2]} \right) + \delta_c(2, l, 3) \quad (522)$$

- Decoding at node 3:

$$R_{2 \rightarrow 13}^{(l)} > I \left(X_2; U_{2 \rightarrow 13, l} \middle| X_3 \mathcal{W}_{[2, l]} \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]} \right) + \delta_{dc}(3, l) \quad (523)$$

$$R_{2 \rightarrow 13}^{(l)} + R_{1 \rightarrow 23}^{(l)} > I \left(X_1 X_2; U_{1 \rightarrow 23, l} U_{2 \rightarrow 13, l} \middle| X_3 \mathcal{W}_{[1, l]} \mathcal{V}_{[13, l, 1]} \mathcal{V}_{[23, l, 2]} \right) + \delta''_{dc}(3, l) \quad (524)$$

$$\hat{R}_{2 \rightarrow 3}^{(l)} < R_{2 \rightarrow 3}^{(l)} + I \left(U_{2 \rightarrow 3, l}; X_3 \mathcal{V}_{[13, l, 3]} \middle| \mathcal{W}_{[3, l]} \mathcal{V}_{[23, l, 2]} \right) - \delta_{dp}(3, l) \quad (525)$$

$$\hat{R}_{1 \rightarrow 3}^{(l)} < R_{1 \rightarrow 3}^{(l)} + I \left(U_{1 \rightarrow 3, l}; X_3 U_{2 \rightarrow 13, l} \mathcal{V}_{[23, l, 3]} \middle| \mathcal{W}_{[2, l]} \mathcal{V}_{[13, l, 1]} \right) - \delta'_{dp}(3, l) \quad (526)$$

$$\begin{aligned} \hat{R}_{1 \rightarrow 3}^{(l)} + \hat{R}_{2 \rightarrow 3}^{(l)} &< R_{1 \rightarrow 3}^{(l)} + R_{2 \rightarrow 3}^{(l)} + I \left(U_{2 \rightarrow 3, l}; X_3 \mathcal{V}_{[13, l, 3]} \middle| \mathcal{W}_{[3, l]} \mathcal{V}_{[23, l, 2]} \right) \\ &+ I \left(U_{1 \rightarrow 3, l}; X_3 U_{2 \rightarrow 13, l} \mathcal{V}_{[23, l, 3]} \middle| \mathcal{W}_{[2, l]} \mathcal{V}_{[13, l, 1]} \right) \\ &- I \left(U_{1 \rightarrow 3, l}; U_{2 \rightarrow 3, l} \middle| \mathcal{W}_{[3, l]} \mathcal{V}_{[23, l, 2]} \mathcal{V}_{[13, l, 1]} X_3 \right) - \delta''_{dp}(3, l) \end{aligned} \quad (527)$$

- Encoding at node 3:

$$\hat{R}_{3 \rightarrow 12}^{(l)} > I \left(X_3; U_{3 \rightarrow 12, l} \middle| \mathcal{W}_{[3, l]} \right) + \delta_c(3, l, 12) \quad (528)$$

$$\hat{R}_{3 \rightarrow 1}^{(l)} > I \left(X_3; U_{3 \rightarrow 1, l} \middle| \mathcal{W}_{[1, l+1]} \mathcal{V}_{[13, l, 3]} \right) + \delta_c(3, l, 1) \quad (529)$$

$$\hat{R}_{3 \rightarrow 2}^{(l)} > I \left(X_3; U_{3 \rightarrow 2, l} \middle| \mathcal{W}_{[1, l+1]} \mathcal{V}_{[23, l, 3]} \right) + \delta_c(3, l, 2) \quad (530)$$

- Decoding at node 1:

$$R_{3 \rightarrow 12}^{(l)} > I \left(X_3; U_{3 \rightarrow 12, l} \middle| X_1 \mathcal{W}_{[3, l]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l, 3]} \right) + \delta_{dc}(1, l) \quad (531)$$

$$R_{3 \rightarrow 12}^{(l)} + R_{2 \rightarrow 13}^{(l)} > I \left(X_2 X_3; U_{2 \rightarrow 13, l} U_{3 \rightarrow 12, l} \middle| X_1 \mathcal{W}_{[2, l]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l, 3]} \right) + \delta''_{dc}(1, l) \quad (532)$$

$$\hat{R}_{3 \rightarrow 1}^{(l)} < R_{3 \rightarrow 1}^{(l)} + I \left(U_{3 \rightarrow 1, l}; X_1 \mathcal{V}_{[12, l+1, 1]} \middle| \mathcal{W}_{[1, l+1]} \mathcal{V}_{[13, l, 3]} \right) - \delta_{dp}(1, l) \quad (533)$$

$$\hat{R}_{2 \rightarrow 1}^{(l)} < R_{2 \rightarrow 1}^{(l)} + I \left(U_{2 \rightarrow 1, l}; X_1 U_{3 \rightarrow 12, l} \mathcal{V}_{[13, l+1, 1]} \middle| \mathcal{W}_{[3, l]} \mathcal{V}_{[12, l, 2]} \right) - \delta'_{dp}(1, l) \quad (534)$$

$$\begin{aligned} \hat{R}_{2 \rightarrow 1}^{(l)} + \hat{R}_{3 \rightarrow 1}^{(l)} &< R_{2 \rightarrow 1}^{(l)} + R_{3 \rightarrow 1}^{(l)} + I \left(U_{3 \rightarrow 1, l}; X_1 \mathcal{V}_{[12, l+1, 1]} \middle| \mathcal{W}_{[1, l+1]} \mathcal{V}_{[13, l, 3]} \right) \\ &+ I \left(U_{2 \rightarrow 1, l}; X_1 U_{3 \rightarrow 12, l} \mathcal{V}_{[13, l+1, 1]} \middle| \mathcal{W}_{[3, l]} \mathcal{V}_{[12, l, 2]} \right) \\ &- I \left(U_{2 \rightarrow 1, l}; U_{3 \rightarrow 1, l} \middle| \mathcal{W}_{[1, l+1]} \mathcal{V}_{[12, l, 2]} \mathcal{V}_{[13, l, 3]} X_1 \right) - \delta''_{dp}(1, l) . \quad (535) \end{aligned}$$

The final private rate equations in Theorem 1 follows from a rather simple Fourier-Motzkin elimination procedure [22].

REFERENCES

- [1] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 471–480, 1973.
- [2] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 1–10, 1976.
- [3] T. Berger, “Multiterminal source coding,” in *The Information Theory Approach to Communications*, G. Longo, Ed. Series CISM Courses and Lectures Springer-Verlag, New York, 1978, vol. 229, pp. 171–231.
- [4] S. Y. Tung, “Multiterminal source coding,” Ph.D. Dissertation, Electrical Engineering, Cornell University, Ithaca, NY, May 1978.
- [5] T. Berger and R. Yeung, “Multiterminal source encoding with one distortion criterion,” *Information Theory, IEEE Transactions on*, vol. 35, no. 2, pp. 228–236, Mar 1989.
- [6] T. Berger, Z. Zhang, and H. Viswanathan, “The ceo problem [multiterminal source coding],” *Information Theory, IEEE Transactions on*, vol. 42, no. 3, pp. 887–902, 1996.
- [7] Y. Oohama, “The rate-distortion function for the quadratic gaussian ceo problem,” *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 1057–1070, 1998.

- [8] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic gaussian two-encoder source-coding problem," *Information Theory, IEEE Transactions on*, vol. 54, no. 5, pp. 1938–1961, May 2008.
- [9] A. Wagner, B. Kelly, and Y. Altug, "Distributed rate-distortion with common components," *Information Theory, IEEE Transactions on*, vol. 57, no. 7, pp. 4035–4057, 2011.
- [10] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727 – 734, Nov. 1985.
- [11] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5240 –5257, Aug. 2011.
- [12] R. Timo, A. Grant, and G. Kramer, "Lossy broadcasting with complementary side information," *Information Theory, IEEE Transactions on*, vol. 59, no. 1, pp. 104–131, Jan 2013.
- [13] A. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 735 – 740, Nov. 1985.
- [14] N. Ma and P. Ishwar, "Interaction strictly improves the wyner-ziv rate-distortion function," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, June 2010, pp. 61–65.
- [15] H. Permuter, Y. Steinberg, and T. Weissman, "Two-way source coding with a helper," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2905–2919, 2010.
- [16] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6180 –6195, Sep. 2011.
- [17] N. Ma, P. Ishwar, and P. Gupta, "Interactive source coding for function computation in collocated networks," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4289 –4305, Jul. 2012.
- [18] L. Sankar and H. Poor, "Distributed estimation in multi-agent networks," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 329–333.
- [19] M. Gastpar, "The wyner-ziv problem with multiple sources," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2762 – 2768, Nov. 2004.
- [20] A. Wyner, "The common information of two dependent random variables," *Information Theory, IEEE Transactions on*, vol. 21, no. 2, pp. 163–179, Mar 1975.
- [21] T. Cover and J. Thomas, *Elements of information theory (2nd Ed)*. Wiley-Interscience, 2006.
- [22] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [23] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [24] P. Piantanida, L. Rey Vega, and A. Hero, "A proof of the generalized markov lemma with countable infinite sources," in *Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on*, July 2014.
- [25] W. Uhlmann, "Vergleich der hypergeometrischen mit der binomial-verteilung," *Metrika*, vol. 10, no. 1, pp. 145–158, 1966.
- [26] A. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *Information Theory, IEEE Transactions on*, vol. 28, no. 6, pp. 828–840, Nov 1982.